

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:22:36 ; Search time 41 Seconds

(without alignments)
847.832 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MKLLLMACIVCAVAFARKRFR.....TAKPAPRPHSPSPLEQANQ 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq.19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	22	Human transp
2	219	100.0	219	22	Human transp
3	219	100.0	219	23	Human secreted
4	219	100.0	219	23	Human secreted
5	219	100.0	219	24	Human transp
6	150	68.5	221	20	Human acid sequenc
7	140	63.9	219	20	Human secreted pro
8	9	4.1	75	22	Drosophila melanog
9	9	4.1	322	17	Leishmania chagasi

10	9	4.1	728	22	ABB65430	Drosophila melanog
11	9	4.1	1874	22	AA876532	Corynebacterium gl
12	9	4.1	2969	22	AA90680	C. glutamicum prote
13	9	3.7		24	ABP96450	H1A class I molecu
14	8	3.7		24	ABP80080	N. gonorrhoeae ami
15	8	3.7	65	21	AA827598	Arabidopsis thalia
16	8	3.7	82	21	AA819327	Arabidopsis thalia
17	8	3.7	94	21	AA819326	Arabidopsis thalia
18	8	3.7	106	21	AA819325	Arabidopsis thalia
19	8	3.7	121	21	AA840893	Human ORF657
20	8	3.7	124	21	AA827837	Arabidopsis thalia
21	8	3.7	129	22	AA653380	Amino acid sequenc
22	8	3.7	132	20	AA934678	Chlamydia pneumoni
23	8	3.7	139	21	AA854857	Arabidopsis thalia
24	8	3.7	145	21	AA832749	Eucalyptus grandis
25	8	3.7	145	21	AA807525	Arabidopsis thalia
26	8	3.7	145	21	AA845122	Arabidopsis thalia
27	8	3.7	145	21	AA845149	Arabidopsis thalia
28	8	3.7	181	21	AA806866	Arabidopsis thalia
29	8	3.7	181	21	AA810820	Arabidopsis thalia
30	8	3.7	185	21	AA854822	Arabidopsis thalia
31	8	3.7	193	21	AA806865	Arabidopsis thalia
32	8	3.7	193	21	AA810819	Arabidopsis thalia
33	8	3.7	205	21	AA806864	Arabidopsis thalia
34	8	3.7	205	21	AA810818	Arabidopsis thalia
35	8	3.7	249	24	ABR40639	Triticum aestivum
36	8	3.7	328	21	AA190267	Protein chimera Q
37	8	3.7	372	22	ABG16669	Novel human diagno
38	8	3.7	412	21	AA90266	Protein chimera Q
39	8	3.7	412	24	ABG73864	L. infantum antige
40	8	3.7	421	22	AB867110	Drosophila melanog
41	8	3.7	428	23	AA850963	Mize methyl Cpg b
42	8	3.7	428	24	ABP70532	Histone deacetylase
43	8	3.7	493	22	AB870327	Drosophila melanog
44	8	3.7	511	22	ABG48687	Human liver peptid
45	8	3.7	511	22	AB828667	Peptide #1318 enco

ALIGNMENTS

RESULT 1	
ID	AA860109 standard; Protein; 219 AA.
AA860109	
AC	AA860109;
XX	
DT	28-MAR-2001 (first entry)
XX	
DE	Human transport protein TPPT-29.
XX	
KW	Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW	neurological disorder; cardiovascular disorder; reproductive disorder;
XX	immune disorder; cancer.
OS	Homo sapiens.
XX	
PN	WO200078953-A2.
XX	
PD	28-DEC-2000.
XX	
PF	16-JUN-2000; 2000WO-US16668.
XX	
PR	17-JUN-1999; 99US-0139923.
XX	
PR	10-AUG-1999; 99US-0148177.
XX	
PR	18-AUG-1999; 99US-0149357.
XX	
PR	28-OCT-1999; 99US-0162287.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandhan O, Burford N;
XX	Baughn MR, Azimzai Y, Lu DM, Au-Young J, Patterson C;

DR WPI: 2001-041424/05.
 DR N-PSDB; AAE27729.
 XX Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 XX
 PS Claim 2, Page 130, 165pp; English.
 XX
 CC The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated TPPTS). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 CC
 XX Sequence 219 AA;
 SQ
 Query Match 100.0%; Score 219; DB 22; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1,1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLMACIVCAFAKRRKPPFISGDDNDGHLHSLNIPYGIKRLPPPLYRRPVNTVP 60
 DB 1 MKLLMACIVCAFAKRRKPPFISGDDNDGHLHSLNIPYGIKRLPPPLYRRPVNTVP 60
 QY 61 SYPGNTYTTGTGSPYFWILTSRGFPYVYHIRGFPPLATQLVNPLPRGSPVPVPSRFFSA 120
 DB 61 SYPGNTYTTGTGSPYFWILTSRGFPYVYHIRGFPPLATQLVNPLPRGSPVPVPSRFFSA 120
 QY 121 AAPAPAPPIAAEPAPAAPIATATVAAEPAPAGAVAAEPAAEPAAEPAAEPAAEPAA 180
 DB 121 AAPAPAPPIAAEPAPAAPIATATVAAEPAPAGAVAAEPAAEPAAEPAAEPAAEPAA 180
 QY 181 EAPVGEPAABEPPAPAPATAPKAPABEPHPSPLEQANQ 219
 DB 181 EAPVGEPAABEPPAPAPATAPKAPABEPHPSPLEQANQ 219
 DB 181 EAPVGEPAABEPPAPAPATAPKAPABEPHPSPLEQANQ 219
 RESULT 2
 AAE27862
 ID AAE27862 standard; Protein; 219 AA.
 XX
 AC AAE27862;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 DE Human zsig63 protein.
 XX
 XX Human; secreted salivary protein; zsig63 protein; host defense protein;
 KW immune modulating factor; antipathogenic; cell-cell signalling molecule;
 KW growth factor; cytokine; growth factor hormone activity; dental caries;
 KW infection; tooth decay; periodontal disease; gastrointestinal disease;
 KW thrush; urinary tract infection; vaginal infection; diabetes; obesity;
 KW anti-inflammatory; chronic tissue damage; lung dysfunction; restenosis;
 KW gene therapy; salivary gland dysfunction; prostate gland dysfunction;
 KW forensic DNA profiling; chondrosarcoma; atherosclerosis; chromosome 4.
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Reptide 1..15
 FT Protein /label= Signal_peptide
 FT /note= "Mature human zsig63 protein"
 FT Region 14..21
 FT /note= "Hydrophilic region"
 FT Domain 16..37
 FT /note= "Domain 1"
 FT Region 17..33
 FT /note= "Antigenic epitope"
 FT Region 24..30
 FT /note= "Hydrophilic region"
 FT Domain 38..126

FT Region /note= "Domain 2"
 FT 66..73
 FT /note= "Antigenic epitope"
 FT Region 103..108
 FT /note= "Antigenic epitope"
 FT Region 124..133
 FT /note= "Repeat 1"
 FT 127..219
 FT /note= "Domain 3"
 FT Region 134..138
 FT /note= "Repeat 2"
 FT Region 139..143
 FT /note= "Repeat 3"
 FT Region 144..148
 FT /note= "Repeat 4"
 FT Region 149..153
 FT /note= "Repeat 5"
 FT Region 154..158
 FT /note= "Repeat 6"
 FT Region 159..163
 FT /note= "Repeat 7"
 FT Region 164..168
 FT /note= "Repeat 8"
 FT Region 169..173
 FT /note= "Repeat 9"
 FT Region 174..178
 FT /note= "Repeat 10"
 FT Region 179..183
 FT /note= "Repeat 11"
 FT Region 184..188
 FT /note= "Repeat 12"
 FT Region 187..192
 FT /note= "Hydrophilic region"
 FT Region 189..193
 FT /note= "Repeat 13"
 FT Region 190..197
 FT /note= "Antigenic epitope"
 FT Region 194..198
 FT /note= "Repeat 14"
 FT Region 199..203
 FT /note= "Repeat 15"
 FT Region 202..215
 FT /note= "Antigenic epitope"
 FT Region 204..208
 FT /note= "Repeat 16"
 XX
 XX US2002090677-A1.
 XX
 XX 11-JUL-2002.
 XX
 PD 03-AUG-2001; 2001US-0923236.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX 17-MAR-2000; 2000US-0527345.
 XX
 PA (ADLER) ADLER D A.
 PA (SHEP) SHEPPARD P O.
 XX
 PI Adler DA, Sheppard PO;
 XX
 XX WPI: 2002-642378/69.
 XX
 DR N-PSDB; AAD45050.
 XX
 PT Novel secreted salivary polypeptide, zsig63, useful as antimicrobial
 PT agent for treating microbial infection, dental caries, periodontal
 PT disease, thrush gastrointestinal disease, and for aiding digestion -
 XX
 PS Claim 10; Page 28; 33pp; English.
 XX
 XX The invention relates to human secreted salivary polypeptide designated
 CC as zsig63 and nucleic acid molecules encoding such polypeptides. zsig63
 CC can be used in detecting agonists and antagonists of its activity, and
 CC is also useful as a host defense polypeptide, immune modulating factor,

Key	Location/Qualifiers
FH	1..15
FT	/note="Fusion protein peptide, specifically claimed"
FT	

FT	Region	1..219	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	9..204	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	14..19	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	16..21	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	16..37	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	16..219	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	24..29	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	24..33	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	25..30	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	38..126	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	66..73	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	103..108	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	109..215	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	124..133	/label= Repeat_1
FT			
FT	Region	127..219	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	134..138	/label= Repeat_2
FT			
FT	Region	139..143	/label= Repeat_3
FT			
FT	Region	144..148	/label= Repeat_4
FT			
FT	Region	149..153	/label= Repeat_5
FT			
FT	Region	154..158	/label= Repeat_6
FT			
FT	Region	159..163	/label= Repeat_7
FT			
FT	Region	164..168	/label= Repeat_8
FT			
FT	Region	169..173	/label= Repeat_9
FT			
FT	Region	174..178	/label= Repeat_10
FT			
FT	Region	179..183	/label= Repeat_11
FT			
FT	Region	184..188	/label= Repeat_12
FT			
FT	Region	187..192	/note= "Antigenic peptide, specifically claimed in claim 15"
FT			
FT	Region	189..193	

FT FT /label= Repeat_13
 FT Region 190..197
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 194..198
 FT /label= Repeat_14
 FT Region 199..203
 FT /label= Repeat_15
 FT 202..215
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 204..208
 FT /label= Repeat_16
 FT US2002061701-A1.
 PD 27-JUN-2002.
 XX 03-AUG-2001; 2001US-0922480.
 PF 17-MAR-1999; 99US-124820P.
 PR 17-MAR-2000; 2000US-0527345.
 XX
 PA (ADLER/) ADLER D A.
 XX (SHEPPARD/) SHEPPARD P O.
 PI Adler DA, Sheppard PO;
 XX WPI; 2002-635468/68.
 DR N-PsDB; ABS52633, ABS52634.
 XX
 PT Novel secreted salivary protein, zsig63 and polynucleotide encoding it
 PT useful for treating microbial infections, inflammatory conditions,
 PT dental caries and lung infections associated with cystic fibrosis
 XX
 PS Claim 10; Page 28; 33pp; English.

CC The present invention relates to a new secreted salivary protein, zsig63.
 CC The invention is useful for detecting in a test sample, the presence of
 CC an antagonist or agonist of zsig63 protein activity. The invention is
 CC also useful as an immunogen for producing an antibody to zsig63
 CC polypeptide, zsig63-cytokine fusion proteins or antibody-cytokine fusion
 CC protein are useful for enhancing in vivo killing of target tissues.
 CC Pharmaceutical composition comprising purified zsig63 polypeptide are
 CC useful in the treatment of conditions associated with pathological
 CC microbes, including bacterial, fungal and viral infections. High
 CC expression of zsig63 in salivary gland suggests that anti-microbial
 CC polypeptides are useful for treatment of dental caries (tooth decay),
 CC periodontal disease, thrush and gastrointestinal disease. Other
 CC applications can be used in urinary tract infections, vaginal infections,
 CC prevention of infection in skin and other epithelial wounds. The
 CC polypeptides can be used to establish normal microflora and protect
 CC against pathogenic colonisation and invasion. The invention is useful
 CC when pro-inflammatory activity is desired. Applications for
 CC such pro-inflammatory activity include the treatment of chronic tissue
 CC damage, particularly in areas having a limited or damaged vascular system
 CC e.g., damage in extremities associated with diabetes. Antagonists to
 CC zsig63 polypeptides may be useful as anti-inflammatory agents. The
 CC invention is useful for the treatment of patients having incompetent
 CC immune system, such as AIDS (acquired immunodeficiency syndrome) patients
 CC or individuals that have undergone chemotherapy, radiation treatment. The
 CC invention is also useful for the treatment of lung infections associated
 CC with cystic fibrosis and its agonists or antagonists are useful for
 CC aiding digestion. The present amino acid sequence represents the human
 CC secreted salivary protein zsig63 of the invention. This sequence is
 CC encoded by the human zsig63 gene located on chromosome 4.
 XX
 SQ Sequence 219 AA;

Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACTIVCAVAFARKRPPPIGEDNDGDGPHLPSINIPYGIKMLPPLPYRREPVNTVP 60
 DB 1 MKLLIMACTIVCAVAFARKRPPPIGEDNDGDGPHLPSINIPYGIKMLPPLPYRREPVNTVP 60
 QY 61 SYPGNTVYTDGILPSYPMILTSPPGFYVYHNGFPPLATOLNPPPLPRGPFPPPSRFFSA 120
 DB 61 SYPGNTVYTDGILPSYPMILTSPPGFYVYHNGFPPLATOLNPPPLPRGPFPPPSRFFSA 120
 QY 121 AAAPAAPPIAAEPAAAAPLTTAPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 DB 121 AAAPAAPPIAAEPAAAAPLTTAPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 QY 181 EAPVGEPAAEPSPAEPATAKPAAPEPHPSPLEQANQ 219
 DB 181 EAPVGEPAAEPSPAEPATAKPAAPEPHPSPLEQANQ 219

RESULT 4
 AAU74536
 ID AAU74536 standard; Protein; 219 AA.
 XX
 AC AAU74536;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human zsig63 polypeptide.
 XX
 KW Human; zsig63; chromosome 4q12-4q13; salivary protein; antimicrobial;
 KW microbial infection; tooth decay; periodontal disease; thrush; emphysema;
 KW gastrointestinal disease; urinary tract infection; vaginal infection;
 KW skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
 KW acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
 KW chronic bronchitis; gene therapy; protein therapy.
 XX
 OS Homo sapiens.
 XX
 PN US6331413-B1.
 XX
 PD 18-DEC-2001.
 XX
 PF 17-MAR-2000; 2000US-0527345.
 XX
 PR 17-MAR-1999; 99US-124820P.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Adler DA, Sheppard PO;
 XX
 DR WPI; 2002-096707/13.
 XX
 DR N-PsDB; AAS20591.
 XX
 PT Polynucleotides encoding salivary proteins useful as anti-microbial
 PT agents -
 XX
 PS Claim 1; Column 49-52; 29pp; English.

CC The invention relates to a polynucleotide derived from the 4q12-4q13
 CC region of human chromosome 4 and encoding a zsig63 polypeptide, a
 CC secreted salivary protein with anti-microbial activity. Due to their
 CC microbial activity, the sequences can be used in the study of microbial
 CC infections, e.g. for recombinant production of anti-microbial proteins.
 CC The sequences can be used in the treatment of tooth decay, periodontal
 CC disease, thrush, gastrointestinal disease, urinary tract infections,
 CC vaginal infections, skin infections, epithelial wounds, chronic tissue
 CC damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung
 CC infections, sarcoidosis, emphysema and chronic bronchitis. This sequence
 CC represents human zsig63.
 XX
 SQ Sequence 219 AA;

Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKLLIMCIVVAPARKRRPFPIGEDNDNDGHPILPSPINIVYGRINLPPPLTYRPVNTVP	60
Db	1	MKLLIMACTVVAAPRRKRRPFPIGEDNDNDGHPILPSPINIVPGIRINLPPPLTYRPVNTVP	60
Qy	61	SYPGNTYDTGLPSYFWILTSRGFPYVYHIGKGFPLATOLANYPPLPRGFPVPYPSRPFSA	120
Db	61	SYPGNTYDTGLPSYFWILTSRGFPYVYHIGKGFPLATOLANYPPLPRGFPVPYPSRPFSA	120
Qy	121	AAAPAPPIAAEPAAAADLTATPVAAEPAGAPVAAEPAAAPVGAEPAAAPVAAEPAA	180
Db	121	AAAPAPPIAAEPAAAADLTATPVAAEPAGAPVAAEPAAAPVGAEPAAAPVAAEPAA	180
Qy	181	EAPVGEPAAEPPSPAPATAKKPAAPRPHSPSPLEQANQ	219
Db	181	EAPVGEPAAEPPSPAPATAKKPAAPRPHSPSPLEQANQ	219
RESULT 5			
ID	ABU08515	standard; Protein; 219 AA.	
XX	ABU08515;		
XX	28-MAY-2003	(first entry)	
XX	Human zsig63 protein.		
XX	Human; immunogen; zsig63; adhesin; salivary gland; dental caries;		
KW	periodontal disease; thrush; gastrointestinal disease; epithelial wound;		
KM	urinary tract infection; vaginal infection; skin infection;		
KM	pro-inflammatory; chronic tissue damage; vascular system; diabetes; AIDS;		
KM	lung infection; cystic fibrosis; lung dysfunction; digestive;		
KM	salivary gland carcinoma; Pneumocystis carinii infection; emphysema;		
KW	chronic bronchitis; prostate dysfunction; prostate adenocarcinoma;		
KM	cell culture media; gene therapy; human chromosome 4q12-4q13;		
KM	dentinogenesis imperfecta; dentin dysplasia type II.		
XX	Homo sapiens.		
OS			
XX	Key	Location/Qualifiers	
FT	Peptide	1..15	
FT		/label= "Signal peptide	
FT		/note= "Specifically claimed in claim 8"	
FT	Protein	16..219	
FT		/label= Mature zsig63	
FT		/note= "Specifically claimed in claim 1"	
FT	Domain	16..37	
FT		/label= Domain_1	
FT		/note= "Specifically claimed in claim 1"	
FT	Domain	38..126	
FT		/label= Domain_2	
FT		/note= "Specifically claimed in claim 1"	
FT	Domain	127..219	
FT		/label= Domain_3	
FT		/note= "Specifically claimed in claim 1"	
FT	Peptide	14..19	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	14..19	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	14..19	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	16..21	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	24..29	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	25..30	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	187..192	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	24..33	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	17..33	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	66..73	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	103..108	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	190..197	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	202..215	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	190..215	
FT		/label= Immunogenic peptide	
FT		/note= "Specifically claimed in claim 15"	
FT	Peptide	124..133	
FT		/label= Repeat_1	
FT		134..138	
FT		/label= Repeat_2	
FT		139..143	
FT		/label= Repeat_3	
FT		144..148	
FT		/label= Repeat_4	
FT		149..153	
FT		/label= Repeat_5	
FT		154..158	
FT		/label= Repeat_6	
FT		159..163	
FT		/label= Repeat_7	
FT		164..168	
FT		/label= Repeat_8	
FT		169..173	
FT		/label= Repeat_9	
FT		174..178	
FT		/label= Repeat_10	
FT		179..183	
FT		/label= Repeat_11	
FT		184..188	
FT		/label= Repeat_12	
FT		189..193	
FT		/label= Repeat_13	
FT		194..198	
FT		/label= Repeat_14	
FT		199..203	
FT		/label= Repeat_15	
FT		204..208	
FT		/label= Repeat_16	
FT	Region	US2002173027-A1.	
FT	Region	21-NOV-2002.	
FT	Region	03-AUG-2001; 2001US-0922469.	
FT	Region	17-MAR-1999; 99US-124820P.	
FT	Region	17-MAR-2000; 2000US-0527345.	
FT	Region	(ADLER/) ADLER D A.	
FT	Region	(SHEP/) SHEPPARD P O.	
FT	Region	Adler DA, She	

DR N-PSDB; ABX93594.

PT Novel isolated zsig63 polypeptide, member of the adhesin family, useful
PT for treating dental caries, periodontal disease, thrush,
PT gastrointestinal disease, urinary tract infections, vaginal infections,
PT skin infections -
XX
PS Claim 10; Page 27-28; 32pp; English.

The invention relates to an isolated zslg63 polypeptide comprising at least 90% identity to an amino acid sequence which comprises domain 1 of zslg63, domain 2, domain 3, mature zslg63 and full length zslg3. Also included are the polynucleotide encoding zslg63, a zslg63 expression vector, a cultured cell comprising the vector and expressing the protein, a DNA encoding a fusion protein (comprising amino acids 1-15, 16-37, 38-126, 127-219 or 16-219 of zslg63 and an additional protein), using a zslg63 reporter gene construct to identify zslg63 agonists, and producing an anti-zslg63 antibody using zslg63 immunogenic peptides, zslg63 is useful for detecting in a test sample, the presence of antagonist of zslg63 protein activity. Zslg63 has antimicrobial activity and since exhibits high expression in salivary gland, can be used for treating dental caries, periodontal disease, thrush, and gastrointestinal disease, urinary tract infections, vaginal infections, skin infections and other epithelial wounds. The polypeptides can be used to establish normal microflora and protect against pathogenic colonization and invasion. Zslg63 can also be used for providing pro-inflammatory activity for treating chronic, tissue damage particularly in areas having limited or damaged vascular system, e.g. in diabetes, and for treating immunocompromised AIDS patients or in individuals that have undergone chemotherapy, radiation treatment, for treating lung infections e.g. in cystic fibrosis. Detection of zslg63 polypeptide at relatively high levels in the tirchea may indicate that such polypeptides may serve as a marker of lung dysfunction. Zslg63 is also useful in diagnosing conditions associated with salivary gland or lung dysfunction including salivary gland carcinoma, Pneumocystis carinii infection, emphysema, chronic bronchitis, prostate dysfunctions such as prostate adenocarcinoma, aiding digestion, and as components of defined cell culture media and may be used to replace serum that is commonly used in culture. The DNA is useful in gene therapy applications to increase or inhibit zslg63 activity, and for detecting abnormalities on human chromosome 4 (e.g. 4q12-4q13) associated with dentinogenesis imperfecta, and dentin dysplasia type II). Zslg63 is an adhesin family member. The present sequence represents human zslg63.

SQ Sequence 219 AA;

Query Match	100.0%	Score 219	DB 24	Length 219
Best Local Similarity	100.0%	Pred. No. 1.1e-200		
Matches 219	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Cy	1	MKLLIMACIVCAFAFKRRFPFPIGEDNDNDGHLHSLNIPYQIRNLPEPLPYRPVNTVP	60
Dd	1	MKLIIIMACIVCAFAFKRRFPFIIGEDNDNDGHLHLSLNIPYQIRNLPEPLPYRPNVTVP	60
Cy	61	SYPGNTYTDTGLSPYPMIITSPOGFVYHIKGFPLATQLNVLPPLPRGPFVFPPPSRFSSA	120
Dd	61	SYPGNTYTDGTLSYPWILTSPOGFVYHIKGFPLATQLNVPLPRGPFVFPPPSRFSSA	120
Cy	121	AAAPAAPPLAAEPAAAAAPLTATPVAABPAAGAPVAAAEPAEAAYGAEPAAEAVVAAEPAA	180
Dd	121	AAAPAAPPLAAEPAAAAAPLTATPVAABPAAGAPVAAAEPAEAAYGAEPAAEAVVAAEPAA	180
Cy	181	EAPVGVEPAAEESPSPAETAKPAADAEHPHSPSELANQ	219
Dd	181	EAPVGVEPAAEESPSPAETAKPAADAEHPHSPSELANQ	219

RESULT 6	
AA19472	standard; Protein; 221 AA
XX	
AC	AA19472;
XX	

DT 14-JUL-1999 (first entry)

DE Amino acid sequence of a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS.

Os Homo sapiens.

PN WO9922243-A1

PD 06-MAY-1999.

PF 23-OCT-1998; 98WO-US22376

PR	24-OCT-1997;	97US-0063387.
PR	24-OCT-1997;	97US-0062784.
PR	24-OCT-1997;	97US-0063088.
PR	24-OCT-1997;	97US-0063089.
PR	24-OCT-1997;	97US-0063090.
PR	24-OCT-1997;	97US-0063091.
PR	24-OCT-1997;	97US-0063092.
PR	24-OCT-1997;	97US-0063097.
PR	24-OCT-1997;	97US-0063098.
PR	24-OCT-1997;	97US-0063099.
PR	24-OCT-1997;	97US-0063100.
PR	24-OCT-1997;	97US-0063101.
PR	24-OCT-1997;	97US-0063109.
PR	24-OCT-1997;	97US-0063110.
PR	24-OCT-1997;	97US-0063111.
PR	24-OCT-1997;	97US-0063118.
PR	24-OCT-1997;	97US-0063386.

PA (HUMA-) HUMAN GENOME SCI INC

PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kayw H, Laffleur DM, Moore PA, Ni J, Olsen HS, Rosen CA
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;

DR WPI; 1999-303069/25.

XX

PT New isolated human genes and the secreted polypeptides they encode
XX
PS Claim 11; Page 401-402; 546pp; English.

The specification describes cDNAs/sequences (AAK51322-X61470) encoding human secreted proteins (AA19442-19590). The polynucleotides and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Pathological conditions can also be diagnosed by determining the amount of the polypeptides in a sample or by determining the presence of mutations in the polynucleotides. Specific uses are described for each of the polynucleotides, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and fetal deficiencies, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's disease, lymphatic disorders, schizophrenia, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, diseases of testes, lung or thymus, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for identifying their binding partners.

SQ Sequence 221 AA;

XX

Query Match 68.5%; Score 150; DB 20; Length 221;
Best Local Similarity 100.0%; Pred. No. 7.4e-135;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACIVCAVAFKRRFPFPGEDDNDGHLPSLNIPIYGINLPPPLYYRPVNTVP 60
DB 1 MKLLMACIVCAVAFKRRFPFPGEDDNDGHLPSLNIPIYGINLPPPLYYRPVNTVP 60
QY 61 SYPGNTYTDGTLPSYPMILTSRGPYVYHIRGFPPLATQLVNPPLPKRGFPVPSRPFSA 120
DB 61 SYPGNTYTDGTLPSYPMILTSRGPYVYHIRGFPPLATQLVNPPLPKRGFPVPSRPFSA 120
QY 121 AAAPAPPIAABPAAAPLTATPVAAEPAA 150
DB 121 AAAPAPPIAABPAAAPLTATPVAAEPAA 150

RESULT 7
AAM30653
ID AAM30653 standard; Protein; 219 AA.
AC AAM30653;
DT 12-APR-1999 (first entry)
DE Human secreted protein clone cp16 1 protein.
KW Human; secreted protein; nutritional activity; cytokine; vaccine;
cell proliferation; differentiation; immune stimulation; suppression;
haematopoiesis regulation; tissue growth; activin; inhibitor; chemotactic;
chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
tumour invasion suppression; tumour inhibition.
OS Homo sapiens.
XX MO9901466-A1.
XX 14-JAN-1999.
XX 01-JUL-1998; 98WO-US13813.
XX 27-OCT-1997; 97US-0958304.
XX 02-JUL-1997; 97US-0887195.
XX (GENY) GENETICS INST INC.
XX PA
XX AGostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
PI Spaulding V, Treacy M;
XX WPI: 1999-105994/09.
XX N-PSDB; AAV80740.
XX
XX New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult testes, adult brain, foetal kidney,
PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
PT potential vaccines
XX
XX Claim 24; Page 71-72; 107pp; English.
XX
XX The present sequence represents a human secreted protein from clone
CC cp16 1, deposited as ATCC 98482. Human secreted protein clone
CC polynucleotides and proteins are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Suggested
CC activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccine) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy.

XX SQ Sequence 219 AA;
Query Match 63.9%; Score 140; DB 20; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.5e-125;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACIVCAVAFKRRFPFPGEDDNDGHLPSLNIPIYGINLPPPLYYRPVNTVP 60
DB 1 MKLLMACIVCAVAFKRRFPFPGEDDNDGHLPSLNIPIYGINLPPPLYYRPVNTVP 60
QY 61 SYPGNTYTDGTLPSYPMILTSRGPYVYHIRGFPPLATQLVNPPLPKRGFPVPSRPFSA 120
DB 61 SYPGNTYTDGTLPSYPMILTSRGPYVYHIRGFPPLATQLVNPPLPKRGFPVPSRPFSA 120
QY 121 AAAPAPPIAABPAAAPLT 140
DB 121 AAAPAPPIAABPAAAPLT 140

RESULT 8
ABB64202
ID ABB64202 standard; Protein; 75 AA.
AC ABB64202;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 19398.
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-656860/75.
XX N-PSDB; ABL08305.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure: SEQ ID NO 19398; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 75 AA;
Query Match 4.1%; Score 9; DB 22; Length 75;
Best Local Similarity 100.0%; Pred. No. 0.91;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAAP 128
 |||||
 45 AAAAPAAP 53

RESULT 9
 AAM03688
 ID AAM03688 standard; Protein; 322 AA.

AC AAM03688;
 XX
 XX
 DT 09-MAR-1997 (first entry)
 XX
 DE Leishmania chagasi acidic ribosomal antigen LcPO.
 XX
 KW Leishmania chagasi; acidic ribosomal antigen; LcPO;
 XX
 KW epitope; K39.
 XX
 OS Leishmania chagasi.
 XX
 FH Key Location/Qualifiers
 FT CDS 30..1202
 FT /*tag= a
 FT
 XX
 PN MO9633414-A2.
 XX
 XX
 PD 24-OCT-1996.
 XX
 PF 19-APR-1996; 96WO-US05472.
 XX
 PR 21-APR-1995; 95US-0428414.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG;
 XX
 DR WPI; 1996-485884/48.
 DR N-PSDB; AAT42164.
 XX
 PT New Leishmania acidic ribosomal P-protein family polypeptide - used
 PT to develop prods. for diagnosis, detection and protection against
 PT Leishmania infections
 XX
 PS Disclosure; Page 29-32; 76pp; English.
 XX
 CC Compounds including polypeptides that contain at least an epitope of
 CC the L. chagasi acidic ribosomal antigen LcPO are useful in a variety
 CC of immunoassays for detecting Leishmania infection. Portions of
 CC LcPO (AAT42164) contg. at least the 17 C-terminal amino acids (AAT42165)
 CC have been found to generate a signal in an ELISA that is equivalent
 CC to that generated by the full length LcPO. A combination
 CC polypeptide may also be used, comprising an LcPO epitope along with
 CC an epitope derived from the Leishmania K39 antigen (AAT42166), pref.
 CC the K39 repeat unit antigen having the sequence given in AAM03690.
 XX
 XX
 SQ Sequence 322 AA;

Query Match 4.1%; Score 9; DB 17; Length 322;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAEPAAPAP 138
 |||||
 289 AAEPAAPAP 297

RESULT 10
 AAB65430
 ID AAB65430 standard; Protein; 728 AA.
 XX
 AC AAB65430;

XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 23082.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL09533.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 23082; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX
 SQ Sequence 728 AA;

Query Match 4.1%; Score 9; DB 22; Length 728;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 SAAAPAPAP 127
 |||||
 418 SAAAPAPAP 426

RESULT 11
 AAB76532
 ID AAB76532 standard; Protein; 1874 AA.
 XX
 AC AAB76532;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MCT protein SEQ ID NO:46.
 XX
 KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering.
 XX
 OS Corynebacterium glutamicum.
 XX

PN WO200100805-A2.
 XX 04-JAN-2001.
 PD
 XX 23-JUN-2000; 2000WO-IB00926.
 PF
 XX 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031563.
 PR 09-JUL-1999; 99DE-1032122.
 PR 09-JUL-1999; 99DE-1032124.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032182.
 PR 09-JUL-1999; 99DE-1032190.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032212.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032927.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040830.
 PR 27-AUG-1999; 99DE-1040831.
 PR 27-AUG-1999; 99DE-1040832.
 PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041395.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BAD) BASF AG.
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Habernauer G;
 XX
 DR WPI; 2001-071486/08.
 DR N-PSDB; AAF67765.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation -
 XX
 PS Claim 20; Page 202-207; 1119pp; English.
 XX
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68083 represent sequencing primers which are used in an
 CC example from the present invention.
 XX
 SQ Sequence 1874 AA;

Query Match 4.1%; Score 9; DB 22; Length 1874;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAEAPVAA 176
 |||||
 DB 568 PAAEAPVAA 576
 RESULT 12
 AAG90680
 ID AAG90680 standard; Protein; 2969 AA.
 XX
 AC AAG90680;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 4434.
 XX
 KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAF65899.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT -
 XX
 PS Claim 17; SEQ ID NO: 4434; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 2969 AA;

Query Match 4.1%; Score 9; DB 22; Length 2969;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 168 PAAEAPVAA 176
 |||||
 DB 1663 PAAEAPVAA 1671
 RESULT 13
 ABP96450
 ID ABP96450 standard; peptide; 9 AA.

AC ABP96450;
XX
XX 27-MAY-2003 (first entry)
DE HLA class I molecule heavy chain optimal peptide #1.
XX
XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;
KW T cell.
OS Homo sapiens.
OS Synthetic.
FN WO2003016905-A2.
PD 27-FEB-2003.
PP 15-AUG-2002; 2002WO-GB03753.
XX 16-AUG-2001; 2001GB-0020042.
PR (AVID-) AVIDEX LTD.
PA
PI Jakobsen BK;
PS MPI, 2003-268366/26.
DR
XX
XX Determining whether a T-cell reacts with a predetermined Major
PT Histocompatibility Complex (MHC) type, comprises contacting a sample of
PT T-cell with MHC type molecules and determining whether the MHC
PT molecules activate the T-cell -
XX
XX Disclosure; Page 17, 56pp; English.

CC The present invention describes a method for determining whether a T-cell
CC reacts with a predetermined major histocompatibility complex (MHC) type,
CC which comprises bringing a sample comprising the T-cell into contact with
CC several molecules of the MHC type, each MHC molecule being complexed with
CC a peptide antigen whose contribution to a T-cell receptor binding the MHC
CC MHC-peptide antigen complex is minimised, and determining whether the MHC
CC molecules cause activation of the T-cell. Also described: (1) a cloning
CC vector encoding an MHC subunit (preferably human) into which a nucleotide
CC encoding a peptide antigen sequence of interest can be inserted such that
CC expression of the vector produces a fusion protein comprising the MHC
CC subunit with the peptide antigen fused into it via a linker sequence;
CC (2) a cell transformed with the above vector; (3) a multivalent class I
CC or II MHC/peptide complex, preferably a multimer, in which the peptide
CC antigen is not recognised in a specific manner by T-cells; and (4) a kit
CC for carrying out the above method, comprising the plurality of MHC
CC molecules cited above. The method is useful in determining whether a
CC T-cell reacts with a predetermined major histocompatibility complex type,
CC and in assessing allo-specific T-cell activity, or in matching transplant
CC and donor patients, as well as in monitoring allorreactive responses
CC following a transplant operation. The present sequence represents a human
CC leukocyte antigen (HLA) class I molecule heavy chain optimal peptide,
CC which is given in the exemplification of the present invention.

XX
XX Sequence 9 AA;

Query Match 3.7%; Score 8; DB 24; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Cy 130 AAEPPAAA 137
|||
Db 1 AAEPPAAA 8

RESULT 14
ID ABP80080 standard; Protein; 59 AA.
XX
XX ABP80080;

[illegible]

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0125264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132556.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157755.

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 3.7%; Score 8; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 130 AAEPA AAA 137
Db 19 AAEPA AAA 26
```

Search completed: January 15, 2004, 15:28:31
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:23:11 ; Search time 18 Seconds

(without alignments)
572.158 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MKLLIMACTIVCAVPAKRRF.....TAKPAAPHPSPSLQANQ 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	9	4.1	322	1	RLA0 LEICH
2	9	4.1	323	1	RLA0 LEICH
3	9	4.1	581	1	HEM1 STRO
4	9	4.1	864	1	NIA TOICA
5	8	3.7	107	1	RLA1 LEIPE
6	8	3.7	110	1	RLA3 SCHPO
7	8	3.7	128	1	RL7 CHLPN
8	8	3.7	353	1	CUP5 GALME
9	8	3.7	396	1	P53 MESAU
10	8	3.7	574	1	ATP2 CHLRE
11	8	3.7	575	1	MSPA TREMA
12	8	3.7	576	1	ECR HELVI
13	8	3.7	1477	1	KELC DROME
14	8	3.7	2470	1	NTC2 MOUSE
15	7	3.2	82	1	ANPA_PSEAM
16	7	3.2	101	1	RLI2 METTL
17	7	3.2	101	1	RLI2 METTL
18	7	3.2	102	1	RLI2 METUA
19	7	3.2	105	1	RLA2 DICI
20	7	3.2	107	1	RLA1 CHLRE
21	7	3.2	109	1	RLA1 TRYCR
22	7	3.2	110	1	RLA2 CRYST
23	7	3.2	111	1	RA2A MAIZE
24	7	3.2	111	1	RLI2 AERPE
25	7	3.2	111	1	RLA1 CAEEL
26	7	3.2	111	1	RLA2 ASPFU
27	7	3.2	112	1	RLA1 DROME
28	7	3.2	112	1	RLI2 TALNI
29	7	3.2	114	1	RLA1 RAT
30	7	3.2	114	1	RLA2 EIMTE
31	7	3.2	120	1	CUI9 ARADI
32	7	3.2	125	1	PSAE SPIOL
33	7	3.2	129	1	DYLA_CHLRE

34	7	3.2	136	1	RL28 HUMAN	P46779 homo sapien
35	7	3.2	136	1	RL28 MOUSE	P41105 mus musculu
36	7	3.2	136	1	SR14_HUMAN	P37108 homo sapien
37	7	3.2	136	1	BCCP_PSEAE	P37799 pseudomona
38	7	3.2	163	1	ATPD_CAEEL	Q09544 caenorhabd
39	7	3.2	168	1	ATPD_HUMAN	P30049 homo sapien
40	7	3.2	199	1	CYCY_RHOCA	Q05389 rhodobacter
41	7	3.2	204	1	RS3A_METUA	Q86062 methanosarc
42	7	3.2	208	1	RS6 MYCGE	P47336 mycoplasma
43	7	3.2	215	1	EF1D_WHEAT	P29546 triticum ae
44	7	3.2	215	1	RS6 MYCPN	P75543 mycoplasma
45	7	3.2	220	1	COAT_CMVSI	Q00467 cymbidium m

ALIGNMENTS

RESULT 1

ID	RLA0 LEICH	STANDARD	PRT	322 AA
AC	P39096			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	60S acidic ribosomal protein P0.			
GN	LCPO.			
OS	Leishmania chagasi.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.			
OX	NCBI_TaxID=44271;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN-ISOLATE MHOM/BR/82/BA-2;			
RX	MEDLINE=9422252; PubMed=7513304;			
RA	Skelky Y.A.W., Benson D.R., Elwasila M., Badaro R., Burns J.M. Jr.,			
RA	Reed S.G.;			
RT	"Antigens shared by leishmania species and Trypanosoma cruzi:			
RT	immunological comparison of the acidic ribosomal P0 proteins."			
RL	Infect. Immun. 62:1643-1651 (1994).			
CC	-1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT			
CC	OF E.COLI PROTEIN L10.			
CC	-1- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH			
CC	DIMERS OF P1 AND P2.			
CC	-1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL; L29300; AAA29263.1;			
DR	InterPro; IPR001813; 60S_ribosomal.			
DR	InterPro; IPR001790; Ribosomal_L10.			
DR	Pfam; PF00428; 60S_Ribosomal; 1.			
DR	Pfam; PF00466; Ribosomal_L10; 1.			
KW	Ribosomal protein; Phosphorylation.			
SQ	SEQUENCE 322 AA; 34594 MW; 2389F125356D26D2 CRC64;			
Query Match	4.1%; Score 9; DB 1; Length 322;			
Best Local Similarity	100.0%; Pred. No. 0.66;			
Matches	9; Conservative			
	0; Mismatches			
	0; Indels			
	0; Gaps			
OY	130 AAEPAAAP 138			
Db	289 AAEPAAAP 297			
RESULT 2				
ID	RLA0 LEICH	STANDARD	PRT	323 AA
AC	P39097			

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 60S acidic ribosomal protein P0.
GN LIPO-A AND LIPO-B.
OS Leishmania infantum.
OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OK NCBI_TaxId=5671;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEM 75;
RX MEDLINE=94088674; PubMed=8264730;
RA Soto M., Requena J.M., Alonso C.;
RT "Isolation, characterization and analysis of the expression of the
RT leishmania ribosomal P0 protein genes.";
RL Mol. Biochem. Parasitol. 61:265-274(1993).
CC -1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF E.COLI PROTEIN L10.
CC -1- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
CC DIMERS OF P1 AND P2.
CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X72714; CAAS1264.1; -
DR EMBL; X72714; CAAS1263.1; -
DR InterPro; IPR001813; 60S_ribosomal.
DR InterPro; IPR001790; Ribosomal_L10.
DR Pfam; PF00428; 60S_ribosomal_1.
DR Pfam; PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 323 AA; 34771 MW; 14967BD7A439D69E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAP 138
DB 290 AAEPAAP 298

RESULT 3
HEM1_STRCO STANDARD; PRT; 581 AA.
ID HEM1_STRCO
AC Q9WX15;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.1-) (GluTR).
GN HEMA OR SCO3319 OR SC68.17C.
OS Streptomyces coelicolor.
OC Streptomyces coelicolor; Actinobacteridae; Actinomycetales;
OC Streptomycetales; Streptomycetaceae; Streptomycetes.
OK NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Laikre L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierzbicki A., Woodward J., Barrell B.G., Parkhill J.,

```

```

RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA(Glu).
CC -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC -1- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL939116; CAB45353.1; -
DR PIR; T36267; T36267.
DR HAMAP; MF_00087; atypical; 1.
DR InterPro; IPR000343; GluTR.
DR InterPro; IPR000594; THLP_domain.
DR Pfam; PF00745; GluTR_dimer; 1.
DR Pfam; PF05201; GluTR_N; 1.
DR Pfam; PF05200; GluTR_NAD_bind; 1.
DR TIGRfam; TIGR01035; hemA; 1.
DR PROSITE; PS00747; GluTR; 1.
KW Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT DOMAIN 292 416 INSERT.
FT ACT SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
FT ACT SITE 99 99 BASE (BY SIMILARITY).
SQ SEQUENCE 581 AA; 60562 MW; D4E256B105AFA037 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 VAAEPAA 162
DB 36 VAAEPAA 44

RESULT 4
NIA_VOLCA STANDARD; PRT; 864 AA.
ID NIA_VOLCA
AC P36841;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nitrate reductase [NADH] (EC 1.7.1.1) (NR).
GN NITR.
OS Volvox carterii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OK NCBI_TaxId=3067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F. Nagatensis / HK10;
RX MEDLINE=93013022; PubMed=1398126;
RA Gruber H., Goetinck S.D., Kirk D.L., Schmitt R.;
RT "The nitrate reductase-encoding gene of Volvox carterii: map location,
RT sequence and induction kinetics.";
RL Gene 120:75-83(1992).
CC -1- FUNCTION: Nitrate reductase is a key enzyme involved in the first
CC step of nitrate assimilation in plants, fungi and bacteria.
CC -1- CATALYTIC ACTIVITY: Nitrite + NAD(+) + H(2)O = nitrate + NADH.
CC -1- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
CC GROUP IS CALLED CYTOCHROME B-557.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- INDUCTION: By nitrate.
CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC N-TERMINAL DOMAIN.

```



```

CC -1- SIMILARITY: Contains 1 cytochrome b5 heme-binding domain.
CC -1- SIMILARITY: TO PAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X64136; CAA45497.1; -.
CC PIR: J01432; J01432.
CC HSSP: P04166; 1B5M.
CC InterPro: IPR001199; Cyt_B5.
CC InterPro: IPR001834; Cyt_B5_reductase.
CC InterPro: IPR000572; Euk_Mb_oxred.
CC InterPro: IPR001709; FPN_cyt_redctse.
CC InterPro: IPR005066; Mo-co_dimer.
CC InterPro: IPR001433; Oxred_PAD/NAD(P).
CC Pfam: PF00970; FAD_binding_6; 1.
CC Pfam: PF00173; heme_1; 1.
CC Pfam: PF03404; Mo-co_dimer; 1.
CC Pfam: PF00175; NAD_binding_1; 1.
CC Pfam: PF00174; oxidored_molb; 1.
CC PRINTS: PR00406; CYTB5RDTASE.
CC PRINTS: PR00363; CYTOCHROME_B5.
CC PRINTS: PR00407; EUMOPTERIN.
CC PRINTS: PR00371; FPNCR.
CC ProDom: PD00612; Cyt_B5; 1.
CC PROSITE: PS00191; CYTOCHROME_B5_1; 1.
CC PROSITE: PS0255; CYTOCHROME_B5_2; 1.
CC PROSITE: PS00559; MOLYBDOPTERIN_EUK; 1.
CC Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
CC Nitate assimilation.
CC METAL: 139 139 MOLYBDENUM-PTERIN (POTENTIAL).
CC METAL: 193 193 MOLYBDENUM-PTERIN (POTENTIAL).
CC DISULFID: 376 376 INTERCHALIN (POTENTIAL).
CC METAL: 532 532 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC METAL: 555 555 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SEQUENCE: 864 AA; 96402 MW; 499529652CDD1C7 CRC64;

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAPP 128
Db 478 AAAAPAPP 486

RESULT 5
RLAL LEIPE STANDARD; PRT; 107 AA.
AC 04613;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 60S acidic ribosomal protein p1.
OS Leishmania peruviana.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
ON NCB1_TaxID=5681;
RX NCB1_TaxID=5681;
RP SEQUENCE FROM N.A.
RC STRAIN=MHOM/PE/84/LC26;
RL de los Santos M., Carrillo C., Panebra A., Montoya Y.;
CC Submitted (JAN-1998) to the EMBL/GenBank/DBS databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

```

```

CC -1- SIMILARITY: Contains 1 cytochrome b5 heme-binding domain.
CC -1- SIMILARITY: TO PAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF045249; AAC02701.1; -.
CC InterPro: IPR001813; 60S_ribosomal.
CC InterPro: IPR001859; Ribosomal_P2.
CC Pfam: PF00428; 60S_ribosomal; 1.
CC PRINTS: PR00456; RIBOSOMALP2.
CC Ribosomal protein.
CC SEQUENCE: 107 AA; 10896 MW; 9117DEC5E490A071 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 126
Db 71 SAAAPAA 78

RESULT 6
RLA3 SCHPO STANDARD; PRT; 110 AA.
AC P17477;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein p1-alpha 3 (A3).
GN RPA3 OR SPBC389.13C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCB1_TaxID=4896;
RX NCB1_TaxID=4896;
RP SEQUENCE FROM N.A.
RC MEDLINE=9020620; PubMed=2325655;
RA Beltrame M., Bianchi M.E.;
RT "A gene family for acidic ribosomal proteins in Schizosaccharomycetes
RT pombe: two essential and two nonessential genes.";
RL Mol. Cell. Biol. 10:2341-2348(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsle K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voiclaert G., Aert R., Robben J., Gymnopre B.,
RA Weltjens I., Vanstrele E., Rieger W., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer B., Moesli D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Dage R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,

```

RA Cernuti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Bartell B.G., Nurse P.;
RA "The genome sequence of *Schizosaccharomyces pombe*.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- MISCELLANEOUS: YEASTS CONTAIN 4 INDIVIDUAL SMALL RIBOSOMAL A
CC PROTEINS (RPA) WHICH CAN BE CLASSIFIED INTO TWO COUPLES OF SIMILAR
CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO
CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.
CC -1- MISCELLANEOUS: RPA3 AND RPA4 ARE ESSENTIAL FOR CELL SURVIVAL,
CC WHEREAS RPA1 AND RPA2 ARE NOT.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M3139; AAA35336.1; -
CC EMBL; AL022070; CAAL7793.1; -
CC PIR; C34715; K6BYP3. -
CC GeneDB; SPombe; SPBC3B9.13c; -
CC InterPro; IPR001813; 608_ribosomal.
CC Pfam; PF00428; 608_ribosomal; 1.
CC Ribosomal protein; Phosphorylation; Multigene family.
SQ SEQUENCE 110 AA; 11171 MW; 173AAC9779F5A91 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 126
DB 64 SAAAPAA 71

RESULT 7
RL7_CHLPN STANDARD; PRT; 128 AA.
ID RL7_CHLPN
AC Q9Z9A1; Q9JQ70;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPL7 OR RL7 OR CPN0080 OR CP0695.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9920606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Greenwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickley E.K., Peterson J., Uettermann T., Berry K., Bass S.,
RA Liner K., Weisman J., Knout H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia*
RT *pneumoniae* AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=0138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatori M., Kohara S., Nakazawa T.;
RT "Comparison of whole genome sequences of *Chlamydia pneumoniae* J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC ACCURATE TRANSLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE001593; AAD18233.1; -
CC EMBL; AE002228; AAF38503.1; -
CC EMBL; AP002545; BAA98290.1; -
CC PIR; C72122; C72122.
CC PIR; H86500; H86500.
CC HSPC; P02392; ICTF.
CC PHCI-2DPAGE; Q9Z9A1; -
CC TIGR; CP0695; -
CC HAMAP; MF 00368; -; 1.
CC InterPro; IPR000206; Ribosomal_L12.
CC Pfam; PF00542; Ribosomal_L12; 1.
CC Prodom; PD001326; Ribosomal_L12; 1.
CC TIGRFAMs; TIGR00855; L12; 1.
CC Ribosomal protein; Complete proteome.
FT INIT MET 0 BY SIMILARITY.
SQ SEQUENCE 128 AA; 13461 MW; 4E2F17A8B057CC CRC64;

Query Match 3.7%; Score 8; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAAP 178
DB 49 EAPVAAAP 56

RESULT 8
CUP5_GALME STANDARD; PRT; 353 AA.
ID CUP5_GALME
AC Q24998;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pupal cuticle protein PCP52 precursor (GMPCC52).
GN PCP52.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditryta; Pyraloidea;
OC Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=epidermis;
RX MEDLINE=95291282; PubMed=7773255;
RA Kollberg U., Obermayer B., Hirsch H., Keilber G., Wolbert P.;
RT "Expression cloning and characterization of a pupal cuticle protein
RT cDNA of *Galleria mellonella* L.";
RL Insect Biochem. Mol. Biol. 25:355-363(1995).
CC -1- FUNCTION: COMPONENT OF THE CUTICLE OF THE PUPA OF GALLERIA
CC MELLONELLA.

```

CC -1- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION DURING THE FIRST DAY AFTER
CC PUPAL ECDYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X77514; CAA54650.1; -
CC PIR; S41958; S41958.
CC Structural protein; Cuticle; Signal.
CC SIGNAL
CC CHAIN 1 15 POTENTIAL.
CC FT 16 353 POLY-ALA.
CC FT DOMAIN 235 243
CC SEQUENCE 353 AA; 35818 MW; C3A35B04EC049172 CRC64;
CC
CC Query Match
CC Best Local Similarity 100.0%; Pred. No. 6.1; Length 353;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Oy 119 SAAAPAA 126
CC |||||
CC 236 SAAAPAA 243
CC
CC RESULT 9
CC PS3_MESAU STANDARD; PRT; 396 AA.
CC ID PS3_MESAU P97276;
CC AC 000366; P97276;
CC DT 01-DEC-1992 (Rel. 24, Created)
CC DT 01-DEC-1992 (Rel. 24, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Cellular tumor antigen p53 (Tumor suppressor p53).
CC GN TP53.
CC OS Mesocricetus auratus (Golden hamster).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
CC OC Mesocricetus.
CC NCBI_TaxID=10036;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Syrian; TISSUE=Kidney;
CC RA MEDLINE=92210007; PubMed=1555773;
CC RX Legros Y., McIntyre P., Soussi T.;
CC RT "The CDNA cloning and immunological characterization of hamster p53.";
CC RL Gene 112:247-250(1992).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RA Hou E.W., Wiseman R.;
CC RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.
CC CC -1- FUNCTION: Acts as a tumor suppressor in many tumor types; induces
CC growth arrest or apoptosis depending on the physiological
CC circumstances and cell type. Involved in cell cycle regulation as
CC a trans-activator that acts to negatively regulate cell division
CC by controlling a set of genes required for this process. One of
CC the activated genes is an inhibitor of cyclin-dependent kinases.
CC Apoptosis induction seems to be mediated either by stimulation of
CC Bax and Fas antigen expression, or by repression of Bcl-2
CC expression.
CC CC -1- SUBUNIT: Binds DNA as a homodimer (By similarity).
CC CC -1- SUBCELLULAR LOCATION: Nuclear.
CC CC -1- DISEASE: p53 is found in increased amounts in a wide variety
CC of transformed cells. p53 is frequently mutated or inactivated
CC in many types of cancer.
CC CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M75144; AAA37085.1; -
CC DR EMBL; U07182; AAB41344.1; -
CC DR PIR; JH0633; JH0633.
CC DR HSSP; P04637; ITUP.
CC DR InterPro; IPR002117; P53.
CC DR Pfam; PF00870; P53; 1.
CC DR PRINTS; PR00386; P53SUPPRESSR.
CC DR PRODOM; PD002681; P53; 1.
CC DR PROSITE; PS00348; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nucleic acid binding; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 45 TRANSCRIPTION ACTIVATION (ACIDIC).
CC FT DNA_BIND 105 295 BY SIMILARITY.
CC FT DOMAIN 328 359 OLIGOMERIZATION.
CC FT DOMAIN 371 390 BASIC (REPRESSION OF DNA-BINDING).
CC FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 15 15 PHOSPHORYLATION (BY PRPK) (BY
CC FT MOD_RES 395 395 SIMILARITY).
CC FT MOD_RES 188 188 PHOSPHORYLATION (BY SIMILARITY).
CC FT CONFLICT 188 188 G -> S (IN REF. 2).
CC SEQUENCE 396 AA; 43631 MW; 906EF0256809BE3 CRC64;
CC
CC Query Match
CC Best Local Similarity 100.0%; Pred. No. 6.7; Length 396;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Oy 120 AAAAPAP 127
CC |||||
CC 66 AAAAPAP 73
CC
CC RESULT 10
CC ATP2_CHLRE STANDARD; PRT; 574 AA.
CC ID ATP2_CHLRE
CC AC P38482;
CC DT 01-OCT-1994 (Rel. 30, Created)
CC DT 01-OCT-1994 (Rel. 30, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14).
CC GN ATP2.
CC OS Chlamydomonas reinhardtii.
CC OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC OC Chlamydomonadaceae; Chlamydomonas.
CC NCBI_TaxID=3055;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA MEDLINE=9253386; PubMed=1386535;
CC RX Franzen L.-G., Falk G.;
CC RT "Nucleotide sequence of cDNA clones encoding the beta subunit of
CC RT mitochondrial ATP synthase from the green alga Chlamydomonas
CC RT reinhardtii: the precursor protein encoded by the cDNA contains both
CC an N-terminal presequence and a C-terminal extension.";
CC RL Plant Mol. Biol. 19:771-780(1992).
CC RN [2]
CC RP STRUCTURE BY NMR OF 1-26.
CC RA MEDLINE=9632639; PubMed=8706917;
CC RX Lancelotti J.-M., Gans P., Bouchayer E., Bally I., Arlaud G.J.,
CC RA Jacquot J.-P.;
CC RT "NMR structures of a mitochondrial transit peptide from the green
CC RT alga Chlamydomonas reinhardtii.";
CC RX FEBS Lett. 391:203-208(1996).
CC CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)

```

```

CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC      -1- SUBCELLULAR LOCATION: Mitochondrial.
CC      -1- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X61624; CAA43808.1; -.
CC      PIR: S23530; S23530.
CC      HSSP: P00829; 1BMF.
CC      InterPro: IPR003593; AAA_ATPase.
CC      InterPro: IPR000793; ATPase_a/bc.
CC      InterPro: IPR000194; ATPase_a/bcentre.
CC      InterPro: IPR004100; ATPase_a/bN.
CC      InterPro: IPR005722; F1_ATPase_beta.
CC      Pfam: PF00006; ATP-synt_ab; 1.
CC      Pfam: PF02874; ATP-synt_ab_C; 1.
CC      SMART: SM00382; AAA; 1.
CC      TIGRfam: TIGR01039; atpd; 1.
CC      PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
CC      ATP synthetase; CF(1); Hydrogen ion transport;
CC      Hydrolyase; ATP-binding; Mitochondrion; Transmembrane;
CC      TRANSIT 1 26 MITOCHONDRION (POTENTIAL).
CC      CHAIN 27 574 ATP SYNTHASE BETA CHAIN.
CC      NP BIND 183 190 ATP (BY SIMILARITY).
CC      SO SEQUENCE 574 AA; 61821 MW; 22B3C6C6D18FBCE CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 574;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      130 AAEPAAAA 137
DB      26 AAEPAAAA 33

RESULT 11
MSPA_TREMA STANDARD; PRT; 575 AA.
ID MSPA_TREMA STANDARD; PRT; 575 AA.
AC Q92413;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Major outer membrane protein mspa precursor (Major sheath protein).
GN MSPA.
OS Treponema maltophilum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_Taxid=51160;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-34.
RC STRAIN=ATCC 51939 / BR;
RA MEDLINE=99121045; PubMed=9922270;
RA Heuner K., Choi B.-K., Schade R., Moter A., Otto A., Goebel U.B.;
RT "Cloning and characterization of a gene (mspa) encoding the major
RT sheath protein of Treponema maltophilum ATCC 51939 (T).";
RL J. Bacteriol. 181:1025-1029 (1999).
RN [2]
RP SUBCELLULAR LOCATION.
RA MEDLINE=2123882; PubMed=1131313;
RA Heuner K., Melzer U., Choi B.-K., Goebel U.B.;
RT "Outer sheath associated proteins of the oral spirochete Treponema
RT maltophilum.";
RL FEWS Microbiol. Lett. 197:187-193 (2001).
CC -1- FUNCTION: Major component of the outer membrane sheath.
CC -1- SUBCELLULAR LOCATION: Outer membrane-associated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y17800; CAA76862.1; -.
CC      KW Outer membrane; Signal.
CC      FT SIGNAL 1 19
CC      CHAIN 20 575
CC      FT CHAIN 20 575 MAJOR OUTER MEMBRANE PROTEIN MSPA.
CC      SO SEQUENCE 575 AA; 62250 MW; C666B188F5DE16A0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 575;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      155 AAEPAAEA 162
DB      19 AAEPAAEA 26

RESULT 12
ECR_HELVI STANDARD; PRT; 576 AA.
ID ECR_HELVI STANDARD; PRT; 576 AA.
AC O18473;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone
DE receptor) (20E receptor) (HVECR).
GN ECR OR NR1H1.
OS Heliothis virescens (Noctuid moth) (Owllet moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Diptera; Noctuoidea;
OC Noctuidae; Heliothinae; Heliothis.
OX NCBI_Taxid=7102;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRC;
RA MEDLINE=99457718; PubMed=10528411;
RA Martinez A., Scanlon D., Gross B., Perara S.C., Palli S.R.,
RA Greenland A.J., Windass J., Pongs O., Broad P., Jepson I.;
RT "Transcriptional activation of the cloned Heliothis virescens
RT (Lepidoptera) ecdysone receptor (HVECR) by muisteroneA.";
RL Insect Biochem. Mol. Biol. 29:915-930 (1999).
CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
CC ELEMENTS (ECRES) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR1
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y09009; CAA70212.1; -.
CC      HSSP: P20393; 1AGY.
CC      InterPro: IPR000536; Hormone_rec_1lg.
CC      InterPro: IPR001723; Steroid_receptor.
CC      InterPro: IPR001628; ZnF_C4steroid.
CC      Pfam: PF00104; hormone_rec; 1.
CC      Pfam: PF00105; zf-C4; 1.
CC      PRINTS: PR00398; STRDHORMONER.
CC      PRINTS: PR00047; STROIDPRNGER.
CC      PRODom: PD000035; ZnF_C4steroid; 1.
CC      SMART: SM00430; HOL1; 1.
CC      SMART: SM00399; ZnF_C4; 1.
CC      PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.

```

KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 FT Zinc-finger. 1 162 MODULATING (POTENTIAL).
 FT DNA BIND 163 228 NUCLEAR RECEPTOR-TYPE.
 FT ZN FING 163 183 C4-TYPE.
 FT ZN FING 199 223 C4-TYPE.
 FT DOMAIN 326 545 HORMONE-BINDING (POTENTIAL).
 SQ SEQUENCE 576 AA: 64638 MW: D13EF787BF263A8 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 576;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 140 TATPVAAE 147
 Db 552 TATPVAAE 559
 RESULT 13
 KELC_DROME STANDARD; PRT; 1477 AA.
 AC Q04652; Q04653; Q9VUA2; Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ring canal kelch protein [contains: Kelch short protein].
 GN KEL OR CG7210.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 CC NCBI_Taxid=7227;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RX MEDLINE=93201592; PubMed=8453663;
 RA Xue F., Cooley L.;
 RT "Kelch encodes a component of intercellular bridges in Drosophila egg
 RT chambers.";
 RL Cell 72:681-693 (1993).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brocktein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostein D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jajala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei J.B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Meruliov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RC CHARACTERIZATION.
 RP TISSUE=Embryo;
 RX MEDLINE=97236487; PubMed=9118811;
 RA Robinson D.N., Cooley L.;
 RT "Examination of the function of two kelch proteins generated by stop
 RT codon suppression.";
 RL Development 124:1405-1417 (1997).
 CC -1- FUNCTION: COMPONENT OF RING CANALS THAT REGULATES THE FLOW OF
 CC CYTOPLASM BETWEEN CELLS. MAY BE INVOLVED IN THE REGULATION OF
 CC CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.
 CC BINDS ACTIN.
 CC -1- SUBCELLULAR LOCATION: INNER SURFACE OF CYTOPLASMIC BRIDGES OR RING
 CC CANALS PRESENT IN EGG CHAMBERS. SUBCORTICALLY IN IMAGINAL DISK
 CC EPITHELIA.
 CC -1- TISSUE SPECIFICITY: BOTH PROTEINS ARE EXPRESSED IN OVARIES, MALE
 CC TESTIS, OVARECTOMIZED FEMALES, CUTICLE, SALIVARY GLAND AND
 CC IMAGINAL DISKS. KELCH ORF1 IS THE PREDOMINANT PROTEIN AND IS ALSO
 CC EXPRESSED IN FAT BODIES. ON ENTRY INTO METAMORPHOSIS LEVELS OF
 CC FULL LENGTH PROTEIN INCREASE IN TESTIS AND IMAGINAL DISKS.
 CC -1- DEVELOPMENTAL STAGE: LARVAE, PUPAE AND ADULTS.
 CC -1- SIMILARITY: Contains 6 BTB/POZ domain.
 CC -1- SIMILARITY: Contains 6 Kelch repeats.
 CC -1- CAUTION: REP.3 BELIEVES RESIDUE 690 IS A SELENOCYSTEINE.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL, L08483; AAAS3471.1; -;
 CC EMBL, L08483; AAAS3472.2; -;
 CC EMBL, AE003657; AAF53651.1; ALT_SEQ.
 CC HSSP; Q05516; IC83.
 CC FLYBase: FBgn0001301; kel.
 CC GO; GO:0007292; P:ogenesis; IMP.
 CC GO; GO:0007301; P:ring canal formation; IDA.
 CC InterPro; IPR000210; BTB_POZ.
 CC InterPro; IPR006651; Kelch.
 CC InterPro; IPR006652; Kelch_rep.
 CC Pfam; PF00651; BTB; 1.
 CC Pfam; PF01344; Kelch; 6.
 CC PRINTS; PR00501; KELCHREPEAT.
 CC SMART; SM00225; BTB; 1.
 CC SMART; SM00612; Kelch; 6.
 CC PROSITE; PS50097; BTB; 1.
 CC Cytoskeleton; Actin-binding; Selenium; Selenocysteine; Kelch repeat;
 KW Repeat.
 KM Repeat.
 FT CHAIN 1 1477 KELCH PROTEIN.
 FT CHAIN 1 689 KELCH SHORT PROTEIN.
 FT DOMAIN 157 223 BTB.
 FT REPEAT 404 449 KELCH 1.
 FT REPEAT 450 496 KELCH 2.
 FT REPEAT 498 543 KELCH 3.
 FT REPEAT 545 592 KELCH 4.
 FT REPEAT 594 639 KELCH 5.
 FT REPEAT 641 687 KELCH 6.
 FT DOMAIN 18 28 ASN-RICH.
 FT DOMAIN 29 87 GLN-RICH.

```

FT DOMAIN 29 36 POLY-GLN.
RA SE_CYS 690 83 POLY-GLN.
FT CONFLICT 493 493 PROBABLE.
FT CONFLICT 596 596 V -> A (IN REF. 1).
FT CONFLICT 824 824 A -> R (IN REF. 1).
FT CONFLICT 858 858 P -> L (IN REF. 1).
FT CONFLICT 1083 1083 G -> D (IN REF. 1).
FT CONFLICT 1086 1083 A -> R (IN REF. 1).
SQ SEQUENCE 1477 AA; 160086 MW; 4851EAE9DDBA47 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 1477;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAP 127
Db 740 AAAAPAP 747

RESULT 14
NTC2 MOUSE STANDARD; PRT; 2470 AA.
AC 03516; 006008; 060941;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
DE B).
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoco Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene."
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178363; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues."
RL Exp. Cell Res. 204:364-372(1993).
[3]
RP SEQUENCE OF 1765-2153 FROM N.A.
RX MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1."
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
[4]
RP FUNCTION.
RX MEDLINE=99396706; PubMed=10393320;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA Tsujimoco Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT embryonic lethality."
RL Development 126:3415-3424(1999).
[5]
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX MEDLINE=95333893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoco Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
RT mouse brain."
RL Brain Res. Mol. Brain Res. 29:263-272(1995).
[6]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

```

```

RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeder E.H., Mumm J.S., Kopan R.;
RT "Murtine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis."
RL J. Biol. Chem. 276:40268-40273(2001).
[7]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(IEC) which are probably linked by disulfide
CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O35516-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O35516-2; Sequence=VSP_001405;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC neuroepithelia, somites, optic vesicles and branchial arches, but
CC not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC the postnatal ependymal cells, and the choroid plexus throughout
CC embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: Contains 35 EGF-like domains.
CC -1- SIMILARITY: Contains 2 Lin/Notch repeats.
CC -1- SIMILARITY: Contains 6 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
CC EMBL; D32210; BAA22094.1; -
CC EMBL; X68279; CAA48340.1; -
CC EMBL; U11881; AAC52924.1; -
CC PIR; A49175; A49175.
CC HSP; P16109; 1FSB.
CC MGD; MGI:97364; Notch2.
CC GO; GO:0005887; C:integral to plasma membrane; IC.
CC GO; GO:0005515; P:protein binding activity; IPI.
CC GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.

```

DR GO; GO:0007219; P:N receptor signaling pathway; IC.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008080; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBLD.
 DR PRINTS; PR00011; EGFAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_RBP_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 22.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KW Receptor; Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 KW Developmental protein; Signal; Phosphorylation;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 KM
 FT SIGNAL 1 25
 FT CHAIN 26 2470
 FT CHAIN 1666 2470
 FT CHAIN 1677 2470
 FT CHAIN 1677 2470
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2470
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 256
 FT DOMAIN 258 294
 FT DOMAIN 296 334
 FT DOMAIN 336 372
 FT DOMAIN 373 411
 FT DOMAIN 413 452
 FT DOMAIN 454 490
 FT DOMAIN 492 528
 FT DOMAIN 530 566
 FT DOMAIN 568 603
 FT DOMAIN 605 641
 FT DOMAIN 643 678
 FT DOMAIN 680 716
 FT DOMAIN 718 753
 FT DOMAIN 755 791
 FT DOMAIN 793 829
 FT DOMAIN 831 869
 FT DOMAIN 871 907
 FT DOMAIN 909 945
 FT DOMAIN 947 983
 FT DOMAIN 985 1021
 FT DOMAIN 1023 1059
 FT DOMAIN 1061 1097
 FT DOMAIN 1099 1145
 FT DOMAIN 1147 1183
 FT DOMAIN 1185 1221
 FT DOMAIN 1223 1260
 FT DOMAIN 1262 1300
 FT DOMAIN 1302 1345
 FT DOMAIN 1347 1382
 FT DOMAIN 1382 1410
 FT DOMAIN 1418 1454
 FT REPEAT 1501 1533
 FT REPEAT 1533 1569
 FT REPEAT 1569 1669

FT REPEAT 1874 1903 ANK 2.
 FT REPEAT 1907 1937 ANK 3.
 FT REPEAT 1941 1970 ANK 4.
 Query Match 3.7%; Score 8; DB 1; Length 2470;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 120 AAAAPAP 127
 DB 2183 AAAAPAP 2190
 RESULT 15
 ANPA_PSEAM STANDARD; PRT; 82 AA.
 ID ANPA_PSEAM
 AC P04002;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antifreeze protein A/B precursor.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Pleuronectiformes;
 OC Pleuronectidae; Pleuronectidae; Pseudopleuronectes.
 OC NCBI_TaxID=8265;
 OK
 RN
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=82197490; PubMed=6952188;
 RA Davies P.L., Roach A.H., Hew C.-L.;
 RT "DNA sequence coding for an antifreeze protein precursor from winter flounder.";
 RT Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
 RL
 RN
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=88259236; PubMed=3133486;
 RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
 RT "Differential amplification of antifreeze protein genes in the pleuronectinae.";
 RT J. Mol. Evol. 27:29-35(1988).
 RL
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem repeats.";
 RT Gene 112:163-170(1992).
 RL
 RN
 RP 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148633; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding mechanism.";
 RT J. Mol. Biol. 223:509-517(1992).
 RL
 RN
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RX MEDLINE=95281060; PubMed=7760940;
 RA Siehler F., Yang D.S.-C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from winter flounder.";
 RT Nature 375:427-431(1995).
 RL
 RN
 RP -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-1 AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

```
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L00138; AAB5964.1; -.
DR      EMBL; L29178; AAB5964.1; JOINED.
DR      EMBL; M62414; AAA49469.1; -.
DR      EMBL; X07506; CAA30389.1; -.
DR      EMBL; M62416; AAA49471.1; -.
DR      EMBL; M62417; AAA49472.1; -.
DR      PIR; A05161; A05161.
DR      PIR; J0704; FDFLAW.
DR      PIR; J0706; J0706.
DR      PIR; S02326; S02326.
DR      PDB; 1A9F; 15-OCT-94.
DR      PDB; 1MPA; 03-JUN-95.
DR      PDB; 1MPB; 03-JUN-95.
DR      InterPro; IPR000104; Antifreeze_1.
DR      PRINTS; PR00308; ANTIFREEZE1.
KW      Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
FT      SIGNAL 1 21
FT      PROPEP 22 44
FT      CHAIN 45 82
FT      VARIANT 36 36
FT      VARIANT 70 70
FT      CONFLICT 24 24
FT      HELIX 45 80
SQ      SEQUENCE 82 AA; 7711 MW; C2AE7B74C0D46CC1 CRC64;
Query Match 3.2%; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      120 AAAAPNA 126
DB      35 AAAAPNA 41
```

Search completed: January 15, 2004, 15:29:01
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:26:21 ; Search time 35 Seconds
(without alignments)
1614.672 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLIMACTVCVAFAPKRRF.....TAKPAAPRHPSPSLQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.1	73	5	Q8MSH9
2	9	4.1	75	5	Q8VBL6
3	9	4.1	344	11	Q9D5J3
4	9	4.1	784	16	Q8PH14
5	9	4.1	841	5	Q9NEA5
6	9	4.1	2969	16	Q8NS46
7	9	3.7	125	10	Q8H4F8
8	8	3.7	139	16	Q8Y2D2
9	8	3.7	141	16	Q8Y2R3
10	8	3.7	145	10	Q9PFC0
11	8	3.7	145	10	Q8E516
12	8	3.7	193	10	Q9C7F7
13	8	3.7	193	10	Q8LE10
14	8	3.7	209	15	Q9DQU8
15	8	3.7	228	16	Q9A4A2
16	8	3.7	231	2	Q93ND2

17	8	3.7	233	2	Q8GGS5	Q8ggs streptomyce
18	8	3.7	242	4	Q8NHG8	Q8nhg homo sapien
19	8	3.7	259	2	Q91UQ4	Q91uq4 rhizobium m
20	8	3.7	262	10	Q8W2Q2	Q8w2q2 oryza sativ
21	8	3.7	266	16	Q8P710	Q8p710 xanthomonas
22	8	3.7	268	16	P95192	P95192 mycobacteri
23	8	3.7	272	10	Q8S2S9	Q8s2s9 thellunglei
24	8	3.7	276	16	Q92WH2	Q92wh2 rhizobium m
25	8	3.7	285	11	Q8BHC2	Q8bhc2 mus musculu
26	8	3.7	291	10	Q40478	Q40478 nicotiana c
27	8	3.7	291	10	Q8H382	Q8h382 oryza sativ
28	8	3.7	302	5	Q9N9A5	Q9n9a5 leishmania
29	8	3.7	325	2	Q929X8	Q929x8 fireaurelia s
30	8	3.7	336	16	Q9AB24	Q9ab24 caulobacter
31	8	3.7	362	11	Q9CX00	Q9cx00 mus musculu
32	8	3.7	364	16	Q8P885	Q8p885 xanthomonas
33	8	3.7	380	11	Q8BZV6	Q8bzv6 mus musculu
34	8	3.7	381	2	Q8E996	Q8e996 acinetobact
35	8	3.7	384	16	Q9AD02	Q9ad02 streptomyce
36	8	3.7	401	8	Q8WEA7	Q8wea7 primula jes
37	8	3.7	428	10	Q941Q8	Q941q8 zea mays (m
38	8	3.7	434	10	Q8H3E5	Q8h3e5 oryza sativ
39	8	3.7	443	8	Q8WEA0	Q8wea0 primula pol
40	8	3.7	448	8	Q31130	Q31130 streptitza
41	8	3.7	468	8	Q9GE00	Q9ge00 primula cor
42	8	3.7	468	8	Q8ME93	Q8me93 primula bul
43	8	3.7	493	2	Q46431	Q46431 klebsiella
44	8	3.7	493	2	Q9W4P4	Q9w4p4 drosophila
45	8	3.7	493	5	Q8S2R1	Q8s2r1 drosophila

ALIGNMENTS

RESULT 1	Q8MSH9	PRELIMINARY:	PRT:	73 AA.
ID	Q8MSH9	01-OCT-2002 (TREMBLrel. 22, Created)		
AC	Q8MSH9	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)		
DE	GH19893P (Fragment).			
GN	BCDNA:GH19893			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkeley;			
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Fries E.,			
RA	George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,			
RA	Miranda A., Mungall C.J., Nunoo U., Paclele J., Paragas V., Park S.,			
RA	Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,			
RA	Celniker S.;			
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY118802; AAM5062.1; -			
DR	PIRbase; P8gn063235; BCDNA:GH19893.			
FT	NON_TER			
SQ	SEQUENCE			
	73 AA; 6847 MW; BC8B77C976694F2D CRC64;			
Qy	Query Match	4.1%; Score 9; DB 5; Length 73;		
	Best Local Similarity	100.0%; Pred. No. 1.1;		
Matches	9; Conservative	0; Mismatches	0; Indels	0; Gaps
DB	120 AAAAPAP 128			
	43 AAAAPAP 51			
RESULT 2				

Q9VBL6 PRELIMINARY; PRT; 75 AA.
 ID Q9VBL6; Q24390;
 AC Q9VBL6; Q24390; (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Accessory gland-specific peptide 57da precursor (Male accessory gland secretory protein 57da).
 GN MST57DA OR CG9074.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 OK NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=OREGON-R;
 RX MEDLINE=95227188; PubMed=7711745;
 RA Stimmerl E., Schaefer M., Schaefer U.,
 RT "Structure and regulation of a gene cluster for male accessory gland transcripts in *Drosophila melanogaster*.";
 RL Insect Biochem. Mol. Biol. 25:127-137(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-O., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
 RA Burkov D., Borchman M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Daya A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jallal M., Kalish F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry B., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venier E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING MATING AND MAY AFFECT EGGLAYING AND BEHAVIOR AFTER MATING.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: LUMEN FLUID OF MALE ACCESSORY GLANDS, BECOMES SEMINAL FLUID.
 CC EMBL; 233647; CAA83925.1; -;
 CC EMBL; A5003753; AAF56515.1; -;
 DR FlyBase; FBgn0011668; Mst57da.
 DR Signal; Behavior.

FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 75 ACCESSORY GLAND-SPECIFIC PEPTIDE 57DA.
 FT DOMAIN 39 64 ALA-PRO-RICH DOMAIN.
 FT CONFLICT 39 46 MISSING (IN REF. 1).
 FT CONFLICT 64 75 MISSING (IN REF. 1).
 SQ SEQUENCE 75 AA; 7106 MW; BCD3D70817C98E14 CRC64;
 Query Match 4.1%; Score 9; DB 5; Length 75;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 120 AAAAPAPP 128
 Db 45 AAAAPAPP 53
 RESULT 3
 ID Q9D5J3 PRELIMINARY; PRT; 344 AA.
 AC Q9D5J3;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE 4930432K09Rik protein.
 GN 4930432K09Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_Taxid=10090;
 OK NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Fukunishi Y., Komu H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Katsukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Koehntgen H.,
 RA Kuehl P., Lewis S., Matsumoto Y., Nishida T., Pesole G., Quackenbush J.,
 RA Schirml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK015291; BAB29782.1; -;
 DR MGD; MGI:1921029; 4930432K09Rik.
 DR InterPro; IPR002965; P-rich extenon.
 DR PRINTS; PR01217; PRICHEXTENON.
 SQ SEQUENCE 344 AA; 36294 MW; D6EAD71CE5802AC CRC64;
 Query Match 4.1%; Score 9; DB 11; Length 344;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MGLLWACT 9
 Db 1 MGLLWACT 9
 RESULT 4
 ID O8PH14 PRELIMINARY; PRT; 784 AA.
 AC O8PH14;
 DT 01-OCT-2002 (Tremblrel. 22, Created)

```

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein XAC3446.
GN XAC3446.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cammaran F., Cardoso J., Chamberg F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Tindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AF011991; AAM38289.1; -
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00486; trans_reg_C_1.
DR ProDom; PD000329; Trans_reg_C_1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 764 AA; 84093 MW; F413770005382EC8 CRC64;

Query Match 4.1%; Score 9; DB 16; Length 784;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 APVAAEPAA 160
DB 160 APVAAEPAA 168

RESULT 5
Q9NEA5 PRELIMINARY; PRT; 841 AA.
AC Q9NEA5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Possible inner dynein arm.
GN L5515.03.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Masny D., Fumelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=9814635; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145 (1998).
DR EMBL; AL161414; CAB77677.1; -.
```

```

DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SMO0320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 841 AA; 93125 MW; BA1B72F1BE821226 CRC64;

Query Match 4.1%; Score 9; DB 5; Length 841;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAAP 127
DB 40 SAAAPAAP 48

RESULT 6
Q8NS46 PRELIMINARY; PRT; 2969 AA.
AC Q8NS46;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE 3-oxoacyl-(acyl-carrier-protein) synthase (EC 2.3.1.85).
GN CGL0836.
OC Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL EMBL; AP005276; BAB98229.1; -
DR EMBL; AP005276; BAB98229.1; -
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR003965; Fatty_acid_synth.
DR InterPro; IPR000794; Ketoacyl_synth.
DR InterPro; IPR002539; Maoc_dehydratase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00698; Acyl_transf_1.
DR Pfam; PF00109; ketoacyl_synth_1.
DR Pfam; PF02801; ketoacyl_synth_C_1.
DR Pfam; PF01575; Maoc_dehydratase; 1.
DR PRINTS; PR01483; FASYNTHASE.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
KW Transferase; Acyltransferase; Complete proteome.
SQ SEQUENCE 2969 AA; 31512 MW; 908B45466D090815 CRC64;

Query Match 4.1%; Score 9; DB 16; Length 2969;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAAPVAA 176
DB 1663 PAAAPVAA 1671

RESULT 7
Q8H4F8 PRELIMINARY; PRT; 125 AA.
AC Q8H4F8;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE OJ1221_H04.16 protein.
GN OJ1221_H04.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

```

OC Eubacteriaceae; Oryzae; Oryza.
OK NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Nipponbare;
RA Sasaki T., Matsunoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
RT clone:O1121.H04."
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004183; BAC21432.1; -.
SQ SEQUENCE 125 AA; 13560 MW; 31F375D8398B8A52 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 AAAPAAP 128
Db 11 AAAPAAP 18

RESULT 8
Q8Y2D2 PRELIMINARY; PRT; 139 AA.
AC Q8Y2D2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical signal peptide protein RSC0404.
GN RSC0404 OR RSC03372.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OK NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,
RA Chandler M., Choienne N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Stigter P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502(2002).
DR EMBL; AL646059; CAD13932.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA; 14153 MW; A3150B6FD629039 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAPAAP 127
Db 55 AAAPAAP 62

RESULT 9
Q8Y2R3 PRELIMINARY; PRT; 141 AA.
AC Q8Y2R3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical protein RSC0272.
GN RSC0272 OR RSC0235.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OK NCBI_TaxID=305;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,
RA Chandler M., Choienne N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Stigter P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502(2002).
DR EMBL; AL646058; CAD13800.1; -.
DR InterPro; IPR001763; Rhodanese-like.
DR SMART; SM00450; RHOD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 15454 MW; 878FE1840A8B8B4E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 AAGAPVAA 156
Db 126 AAGAPVAA 133

RESULT 10
Q9PFC0 PRELIMINARY; PRT; 145 AA.
AC Q9PFC0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Histone H2B like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OK NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones."
RT DNA Res. 4:215-230(1997).
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
DR EMBL; AB005243; BAB10609.1; -.
DR InterPro; IPR004822; Histone core.
DR InterPro; IPR000558; Histone_H2B.
DR Pfam; PF00125; histone_1.
DR PRINTS; PR00621; HISTONEH2B.
DR ProDom; PD000497; Histone_H2B; 1.
DR SMART; SM00427; H2B; 1.
DR PROSITE; PS00357; HISTONE_H2B; 1.
KW Chromosomal protein; DNA-binding; Nuclear protein; Nucleosome core.
SQ SEQUENCE 145 AA; 15732 MW; C68421B559D42B30 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAAPAAA 137
Db 19 AAAPAAA 26

```

```

RESULT 11
ID 096516 PRELIMINARY; PRT; 145 AA.
AC 096516;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Histone H2B like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
RX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ecotype columbia;
RA Phillips G.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC EMBL: Y07745; CAA69025.1; -
CC InterPro: IPR004822; Histone_core.
CC InterPro: IPR000558; Histone_H2B.
CC Pfam: PF00125; histone_1.
CC PRINTS: PR00621; HISTONEH2B.
CC PRODOM: PD000497; HISTONE_H2B.
CC SMART: SM00427; H2B.1.
DR PROSITE: PS00357; HISTONE_H2B; 1.
DR SMART: PS00357; HISTONE_H2B; 1.
DR PROSITE: PS00357; HISTONE_H2B; 1.
DR Chromosomal protein; DNA-binding; Nuclear protein; Nucleosome core.
KM SEQUENCE 145 AA; 15733 MW; 6C8E8B18390F4686 CRC64;
SQ

```

```

Query Match 3.7%; Score 8; DB 10; Length 145;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 130 AAAPAAA 137
Db 19 AAAPAAA 26

```

```

RESULT 12
ID 09C7F7 PRELIMINARY; PRT; 193 AA.
AC 09C7F7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Lipid transfer protein, putative.
GN F1K9.6 OR ATIG27950.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
RX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ec. Columbia;
RA MEDLINE=21016719; PubMed=1130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Alami H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Chou R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Mafti R., Marziani A.,
RA Miltsecher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

```

```

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shin P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC069471; AAG51485.1; -
DR EMBL: AY092956; AAM12955.1; -
DR EMBL: AY128712; AAM91112.1; -
DR InterPro: IPR003612; AA1.
DR Pfam: PF00234; tryp_alpha_amy1.1.
DR SEQUENCE 193 AA; 19759 MW; D54B38B12FEB610 CRC64;
SQ

```

```

Query Match 3.7%; Score 8; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 120 AAAPAP 127
Db 20 AAAPAP 27

```

```

RESULT 13
ID 08LE10 PRELIMINARY; PRT; 193 AA.
AC 08LE10;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Lipid transfer protein, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
RX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzer S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation."
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY085407; AAM62634.1; -
DR InterPro: IPR003612; AA1.
DR Pfam: PF00234; tryp_alpha_amy1.1.
DR SMART: SM00499; AA1.1.
DR SEQUENCE 193 AA; 19747 MW; C091074205D76610 CRC64;
SQ

```

Query Match 3.7%; Score 8; DB 10; Length 193;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAP 127
 |||||
 Db 20 AAAAPAP 27

RESULT 14

Q9DQ08 PRELIMINARY; PRT; 209 AA.
 AC Q9DQ08: 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
 DE 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
 DE Nef protein (Negative factor) (27 kDa protein).
 GN NEF.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=01u09;
 RX MEDLINE=21002575; PubMed=1118071;
 RA Geflin R., Wolf D., Muller R., Hill M.D., Steilweg E., Freitag M.,
 Sasse G., Scott G.B., Baur A.S.;
 RT "Functional and structural defects in HIV-1 nef genes derived from
 pediatric long-term survivors.";
 RL AIDS Res. Hum. Retroviruses 16:1855-1868(2000).
 CC -1- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
 ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4 (T4) ANTIGEN (BY
 SIMILARITY).
 CC EMBL; AF252901; AAG34573.1; -.
 DR HSSP; P03406; IEFN.
 DR InterPro: IPR001558; HIV_Nef.
 DR Pfam; PF00469; F-protein_1.
 DR ProDom; PD000031; HIV_Nef_1.
 KW AIDS; GTP-binding; Lipoprotein; Myristate.
 SQ SEQUENCE 209 AA; 23656 MW; 6B858BAD98C79519 CRC64;

Query Match 3.7%; Score 8; DB 15; Length 209;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 ABPAPAP 138
 |||||
 Db 23 ABPAPAP 30

RESULT 15

Q9A4A2 PRELIMINARY; PRT; 228 AA.
 AC Q9A4A2: 01-JUN-2001 (TRENBLREL. 17, Created)
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
 DE 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Cytochrome c, membrane-bound.
 GN CC2935.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 Kolonay J.F., Smit J., Craven M.B., Kouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Yamathevan J., Ermolaeva M., White O.,
 RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005958; AAK24897.1; -.
 DR HSSP; P81459; 1154.
 DR TIGR; CC2935; -.
 DR InterPro: IPR000104; Antifreeze 1.
 DR InterPro: IPR000345; Cyt_c_heme_bind.
 DR InterPro: IPR003088; Cyt_Ct.
 DR InterPro: IPR002327; Cyt_CTAB.
 DR Pfam; PF00034; Cytochrome_c; 1.
 DR PRINTS; PR00308; ANTIFREEZE1.
 DR PROSITE; PR00604; CYTCHROMETAB.
 DR PROSITE; PR000375; Cyt_CTAB; 1.
 DR PROSITE; PR00190; CYTOCHROME_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 228 AA; 23023 MW; 5AE05C7DD5D05F82 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 228;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAP 127
 |||||
 Db 219 AAAAPAP 226

Search completed: January 15, 2004, 15:29:49
 Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:26:56 : Search time 21 Seconds
(without alignments)
1002.902 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLIMACTVCVAFARKRRF.....TAKPAAPRHPSPSLQANQ 219

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	4.1	581	2 T36267	probable glutamyl-nitrate reductase
2	9	4.1	864	1 JC1422	60S acidic ribosome L7/L12 ribosomal protein
3	8	3.7	110	1 R6BYP3	probable lipid transfer protein
4	8	3.7	129	2 H86500	cytochrome c, membrane
5	8	3.7	129	2 C72122	hypothetical protein
6	8	3.7	193	2 H86404	probable ABC transporter
7	8	3.7	228	2 E87612	DNA binding protein
8	8	3.7	268	2 G70645	electron transfer
9	8	3.7	276	2 G95887	TPR domain protein
10	8	3.7	291	2 T02434	pupal cuticle protein
11	8	3.7	325	2 T48870	phenoxylbenzoate diester
12	8	3.7	336	2 A87300	cellular tumor antigen
13	8	3.7	353	2 A41958	probable farnesyl transferase
14	8	3.7	381	2 T46827	transcription enhancer
15	8	3.7	396	1 JH0633	hypothetical protein
16	8	3.7	495	2 T52066	H+-transporting ATPase
17	8	3.7	506	2 A40679	probable cell division
18	8	3.7	523	2 B40679	GRP-binding protein
19	8	3.7	550	2 T55557	kelch protein, long
20	8	3.7	574	2 S23530	hypothetical protein
21	8	3.7	611	2 C96030	phycocyanin
22	8	3.7	846	2 S52418	hypothetical protein
23	8	3.7	1476	2 A45773	hypothetical protein
24	8	3.7	4307	2 T20721	hypothetical protein
25	7	3.2	38	2 S23173	phycocyanin
26	7	3.2	78	2 E83446	hypothetical protein
27	7	3.2	82	1 FDF1AW	antifreeze protein
28	7	3.2	82	2 S02326	antifreeze protein
29	7	3.2	82	2 A05161	antifreeze protein

30	7	3.2	82	2 I51125	antifreeze protein
31	7	3.2	100	2 F95852	hypothetical protein
32	7	3.2	101	2 H69091	ribosomal protein
33	7	3.2	102	2 D64363	ribosomal protein
34	7	3.2	106	1 R6D0P2	acidic ribosomal protein
35	7	3.2	107	1 R6KMLC	acidic ribosomal protein
36	7	3.2	109	1 R6UTP1	acidic ribosomal protein
37	7	3.2	110	2 T37490	ribosomal protein
38	7	3.2	111	2 E72524	probable ribosomal protein
39	7	3.2	111	2 B95857	hypothetical protein
40	7	3.2	112	1 R5FFZ2	acidic ribosomal protein
41	7	3.2	112	2 S54179	acidic ribosomal protein
42	7	3.2	112	2 E86141	protein T25K16.9
43	7	3.2	114	1 R5RT12	acidic ribosomal protein
44	7	3.2	114	2 R5MS2H	ribosomal protein
45	7	3.2	114	2 F84266	50S ribosomal protein

ALIGNMENTS

RESULT 1
T36267
probable glutamyl-tRNA reductase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36267
R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, June 1999
A:Reference number: 221576
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-581 <MDR>
A:Cross-references: EMBL:AL079345; PIDN:CAB45353.1; GSPDB:GN00070; SCOPDB:SCE68.17C
A:Experimental source: strain A3(2)
A:Genetics:
A:Gene: SCOPDB:SCE68.17C

Query Match 4.1%; Score 9; DB 2; Length 581;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 VAAEPAA 162
DB 36 VAAEPAA 44

RESULT 2
JC1422
nitrate reductase (NADH) (EC 1.7.1.1) - Volvox carterii
C:Species: Volvox carterii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
C:Accession: JC1422; S22192
R:Gruber, H.; Goetinck, S.D.; Kirk, D.L.; Schmitt, R.
Gene 120, 75-83, 1992
A:Title: The nitrate reductase-encoding gene of Volvox carterii: Map location, sequence
A:Reference number: JC1422; MUID:93013022; PMID:1398126
A:Accession: JC1422
A:Molecule type: DNA
A:Residues: 1-864 <GR2>
A:Cross-references: EMBL:X64136; NID:g21993; PIDN:CAA45497.1; PID:g21994
A:Note: Submitted to the EMBL Data Library, January 1992
C:Genetics:
A:Gene: nita
A:Map position: linkage group IX
A:Insertions: 183/3; 234/2; 295/1; 339/1; 372/1; 425/2; 521/3; 593/3; 677/3; 797/2
C:Function: catalyzes the reduction of nitrate to nitrite
A:Description: catalyzes the reduction of nitrate to nitrite
C:Superfamily: nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 reductase
C:Keywords: dimer; electron transfer; PAD; flavoprotein; heme; iron; metalloprotein; mol
F:36-425/Domain: molybdopterin-binding domain homology <PCO>
F:497-571/Domain: cytochrome b5 core homology <CB5>

F:613-864/Domain: cytochrome-b5 reductase homology <CBR>
 F:139/Binding site: molybdopterin (Cys) (covalent) #status predicted
 F:532,555/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 4.1% Score 9; DB 1; Length 864;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
 DB 478 AAAAPAAP 486

RESULT 3

60s acidic ribosomal protein p1-alpha - fission yeast (Schizosaccharomyces pombe)
 N:Alternates names: acidic ribosomal protein 3; ribosomal protein YL2611
 C:Species: Schizosaccharomyces pombe
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 10-Dec-1999

C:Accession: C34715; T40352
 R:Beltrame, M.; Bianchi, M.E.
 Mol. Cell. Biol. 10, 2341-2348, 1990

A:Title: A gene family for acidic ribosomal proteins in Schizosaccharomyces pombe: two
 A:Reference number: A34715; MUID:90220620; PMID:2325655

A:Accession: C34715
 A:Molecule type: DNA

A:Residues: 1-110 <BEL>
 A:Cross-references: EMBL:A333139; NID:g173467; PIDD:AAA5336.1; PID:g173468
 R:Wood, V.; Rajandream, M.A.; Barrett, B.G.; Skelton, J.; Churcher, C.M.
 submitted to the EMBL Data Library, March 1997

A:Reference number: Z21922
 A:Accession: T40352

A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA

A:Residues: 1-110 <WOO>
 A:Cross-references: EMBL:AL022070; PIDD:CAA17793.1; GSPDB:GN00067; SPDB:SPBC3B9.13c
 A:Experimental source: strain 972h-; cosmid c3B9

C:Genetics:
 A:Gene: rpa3
 A:Map position: 2
 A:introns: 23/3; 103/3
 C:Superfamily: rat acidic ribosomal protein P1
 C:Keywords: phosphoprotein; protein biosynthesis; ribosome

Query Match 3.7% Score 8; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 4.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 126
 DB 64 SAAAPAA 71

RESULT 4

H86500
 L7/L12 ribosomal protein [imported] - Chlamydia pneumoniae (strain J138)
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: H86500
 R:Shital, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
 Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
 A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: H86500
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-129 <STO>
 A:Cross-references: GB:BA000008; NID:g8978453; PIDD:BA98290.1; GSPDB:GN00142
 A:Experimental source: strain J138

C:Genetics:
 A:Gene: r17
 C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 3.7% Score 8; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAEP 178
 DB 50 EAPVAAEP 57

RESULT 5

C72122
 ribosomal protein L7/L12 CP0695 [imported] - Chlamydia pneumoniae (strains CWL029 an
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: C72122; G81548
 R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:9920606; PMID:10192388

A:Accession: C72122
 A:Molecule type: DNA

A:Residues: 1-129 <ARN>
 A:Cross-references: GB:AE001593; GB:AE001363; NID:g4376334; PIDD:AAD18233.1; PID:g437633
 A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: G81548
 A:Molecule type: DNA

A:Residues: 1-129 <REA>
 A:Cross-references: GB:AE002228; GB:AE002161; NID:g7189606; PIDD:AAF8503.1; PID:g718960
 A:Experimental source: strain AR39, HL cells

C:Genetics:
 A:Gene: r17; CP0695
 C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 3.7% Score 8; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAEP 178
 DB 50 EAPVAAEP 57

RESULT 6

H86404
 probable lipid transfer protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: H86404
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, R.;
 Nature 408, 816-820, 2000

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H86404
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-193 <STO>
 A:Cross-references: GB:AE005172; NID:g11024871; PIDD:AAQ2695.1; GSPDB:GN00141
 A:Experimental source: strain J138

C:Genetics:
 A:Map position: 1

Query Match 3.7% Score 8; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAP 127
|||||
Db 20 AAAAPAP 27

RESULT 7

cytochrome c, membrane-bound [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 18-Nov-2002

C:Accession: B67612

R:Finan, M.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Land, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:1125647

A:Accession: B67612

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-228 <STO>

A:Cross-references: GB:AB005673; NID:g13424561; PIDN:AAK24897.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2935

C:Superfamily: membrane-bound cytochrome c/cytochrome c homology

C:Keywords: chromoprotein; heme; iron; metalloprotein

F:81.84/Binding site: heme iron (His) (axial ligand) #status predicted

F:150/Binding site: heme iron (Met) (axial ligand) #status predicted

Query Match 3.7%; Score 8; DB 2; Length 228;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAP 127
|||||
Db 219 AAAAPAP 226

RESULT 8

hypothetical protein RV3134c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70645

R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295587; PMID:9634230

A:Accession: G70645

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-268 <COU>

A:Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAM06280.1; PID:e230951;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3134c

Query Match 3.7%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 PSPAPAP 200
|||||
Db 138 PSPAPAP 145

RESULT 9

G95887

probable ABC transporter permease protein Smb20381 [imported] - Sinorhizobium meliloti {

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 01-Mar-2002

C:Accession: G95887

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chalm, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,663-kb pSymB megaplasmid from the N₂-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95887

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-276 <KRP>

A:Cross-references: GB:AL591985; PIDN:CAC48767.1; PID:g15140240; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chalm, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Smb20381

A:Genome: plasmid

C:Superfamily: spermidine/putrescine transport system permease protein potI

Query Match 3.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAEPAAA 137
|||||
Db 268 AAEPAAA 275

RESULT 10

T02434

DNA binding protein EREBP-4 - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 31-Mar-2001

C:Accession: T02434

R:Ohme-Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene response

A:Reference number: Z14671; MUID:95276459; PMID:7756828

A:Accession: T02434

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-291 <OHM>

A:Cross-references: EMBL:D38125; NID:g790361; PIDN:BA07323.1; PID:g1208497

A:Experimental source: strain BY4; tissue-type leaf

C:Superfamily: Arabidopsis thaliana hypothetical protein T19P19.170

Query Match 3.7%; Score 8; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 PAAAPLT 140
|||||
Db 247 PAAAPLT 254

RESULT 11

T48873

electron transfer protein [imported] - Frateuria sp. (strain ANA-18)

C:Species: Frateuria sp.

A:Variety: strain ANA-18

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000

C:Accession: T48873

R.Murakami, S.; Takashima, A.; Takemoto, J.; Takenaka, S.; Shinke, R.; Aoki, K.

Gene 226, 189-198, 1999

A>Title: Cloning and sequence analysis of two catechol-degrading gene clusters from the

A/Reference number: Z24833

A/Accession: T46873

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-325 <MUR>

A/Cross-references: EMBL:AB009373; PIDN:BAW75213.1

A/Experimental source: strain ANA-18

C/Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 325;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAEPAAA 137
Db 223 AAEPAAA 230

RESULT 12

AA87300

TPR domain protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C/Accession: A87300

R/Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A>Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11255647

A/Accession: A87300

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-336 <STO>

A/Cross-references: GB:AE005673; NID:g13421571; PIDN:AAK22397.1; GSPDB:GN00148

C/Genetics:

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 336;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAPAP 127
Db 112 AAAPAP 119

RESULT 13

S41958

pupal cuticle protein precursor - greater wax moth

C/Species: Galleria mellonella (greater wax moth)

C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 19-May-2000

C/Accession: S41958

R/Kollberg, U.; Obermeier, B.; Hirsch, H.; Kelber, G.; Wolbert, P.

submitted to the EMBL Data Library, February 1994

A/Description: Expression cloning and characterization of a pupal cuticle protein cDNA G

A/Reference number: S41958

A/Accession: S41958

A/Molecule type: mRNA

A/Residues: 1-353 <KOL>

A/Cross-references: EMBL:X77514; NID:g453403; PIDN:CAA54650.1; PID:g453404

F/1-15/Domain: signal sequence #status predicted <SIG>

F/16-353/Product: pupal cuticle protein #status predicted <MAT>

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 353;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 SAAAPAA 126
|||||||

Db 236 SAAAPAA 243

RESULT 14

T46827

phenoxybenzoate dioxygenase [imported] - Acinetobacter lwoffii

C/Species: Acinetobacter lwoffii

C/Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000

C/Accession: T46827

R/Kim, S.I.; Leem, S.H.; Choi, J.S.; Chung, Y.H.; Kim, S.; Park, Y.M.; Ha, K.S.

submitted to the EMBL Data Library, November 1996

A/Reference number: Z24100

A/Accession: T46827

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-381 <KIM>

A/Cross-references: EMBL:U7659; PIDN:AAC31770.1

A/Experimental source: strain K24

C/Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 381;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAEPAAA 137
Db 279 AAEPAAA 286

RESULT 15

JH0633

cellular tumor antigen p53 - golden hamster

N/Alternate names: tumor-suppressor protein p53

C/Species: Mesocricetus auratus (golden hamster)

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C/Accession: JH0633

R/Legros, Y.; McIntyre, P.; Soussi, T.

Gene 112, 247-250, 1992

A>Title: The cDNA cloning and immunological characterization of hamster p53.

A/Reference number: JH0633; MUID:9221007; PMID:1555773

A/Accession: JH0633

A/Molecule type: mRNA

A/Residues: 1-396 <LEG>

A/Cross-references: GB:M75144; NID:g191414; PIDN:AA37085.1; PID:g191415

A/Experimental source: kidney, strain MPI

C/Genetics:

A/Genes: p53

C/Superfamily: cellular tumor antigen p53

C/Keywords: apoptosis; cell division control; DNA binding; homotetramer; nucleus; phospho

F/179,182,241,245/Binding site: zinc (Cys, His, Cys, Cys) #status predicted

F/395/Binding site: phosphoryl-RNA (Ser) (covalent) #status predicted

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 396;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAPAP 127
Db 66 AAAPAP 73

Search completed: January 15, 2004, 15:30:23
Job time : 21 secs

DR WPI: 2001-041424/05.
 DR N-PSDB; AAF27729.
 PT Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 XX
 PS Claim 2; Page 130; 165pp; English.
 CC The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated rPTs). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 1178; DB 22; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2, 5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLIACIVCAVFAKRRPPFAGEDNDGHPHPSLNIPIGIRNLPPPLYRPNVTVP 60
 DB 1 MKLLIACIVCAVFAKRRPPFAGEDNDGHPHPSLNIPIGIRNLPPPLYRPNVTVP 60
 QY 61 SYPGNTYDTGCLPSYFWIITSGFPYVYHGRGPLATQNLVPLPRGPFVPSRFFSA 120
 DB 61 SYPGNTYDTGCLPSYFWIITSGFPYVYHGRGPLATQNLVPLPRGPFVPSRFFSA 120
 QY 121 AAAPAPPIAAEPAAAPLTATPVAAPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180
 DB 121 AAAPAPPIAAEPAAAPLTATPVAAPAAAGVAAEPAAEPVGAEPAAEPVAAEPAA 180
 QY 181 EAPVGEPAAEPSPAEPATAPKAPAPRPPSPLEQANQ 219
 DB 181 EAPVGEPAAEPSPAEPATAPKAPAPRPPSPLEQANQ 219
 RESULT 2
 AAE27862
 ID .AAE27862 standard; Protein; 219 AA.
 AC AAE27862;
 DT 27-DEC-2002 (first entry)
 DE Human zsig63 protein.
 DE Human zsig63 protein.
 KW Human; secreted salivary protein; zsig63 protein; host defense protein;
 KW immune modulating factor; antipathogenic; cell-cell signalling molecule;
 KW growth factor; cytokine; growth factor hormone activity; dental carrier;
 KW infection; tooth decay; periodontal disease; gastrointestinal disease;
 KW thrush; urinary tract infection; vaginal infection; diabetes; obesity;
 KW anti-inflammatory; chronic tissue damage; lung dysfunction; restenosis;
 KW gene therapy; salivary gland dysfunction; prostate gland dysfunction;
 KW forensic DNA profiling; chondrosarcoma; atherosclerosis; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..15
 FT Peptide /label= Signal_peptide
 FT Protein /note= "Mature human zsig63 protein"
 FT Region 14..21
 FT /note= "Hydrophilic region"
 FT Domain 16..37
 FT /note= "Domain 1"
 FT Region 17..33
 FT /note= "Antigenic epitope"
 FT Region 24..30
 FT /note= "Hydrophilic region"
 FT Domain 38..126

FT /note= "Domain 2"
 FT 66..73
 FT /note= "Antigenic epitope"
 FT Region 103..108
 FT /note= "Antigenic epitope"
 FT Region 124..133
 FT /note= "Repeat 1"
 FT 127..219
 FT /note= "Domain 3"
 FT Region 134..138
 FT /note= "Repeat 2"
 FT Region 139..143
 FT /note= "Repeat 3"
 FT Region 144..148
 FT /note= "Repeat 4"
 FT Region 149..153
 FT /note= "Repeat 5"
 FT Region 154..158
 FT /note= "Repeat 6"
 FT Region 159..163
 FT /note= "Repeat 7"
 FT Region 164..168
 FT /note= "Repeat 8"
 FT Region 169..173
 FT /note= "Repeat 9"
 FT Region 174..178
 FT /note= "Repeat 10"
 FT Region 179..183
 FT /note= "Repeat 11"
 FT Region 184..188
 FT /note= "Repeat 12"
 FT Region 187..192
 FT /note= "Hydrophilic region"
 FT Region 189..193
 FT /note= "Repeat 13"
 FT Region 190..197
 FT /note= "Antigenic epitope"
 FT Region 194..198
 FT /note= "Repeat 14"
 FT Region 199..203
 FT /note= "Repeat 15"
 FT Region 202..215
 FT /note= "Antigenic epitope"
 FT Region 204..208
 FT /note= "Repeat 16"
 XX
 XX US2002090677-A1.
 XX PN
 XX PD 11-JUL-2002.
 XX PF 03-AUG-2001; 2001US-0923236.
 XX PR 17-MAR-1999; 99US-124820P.
 XX PR 17-MAR-2000; 2000US-0527345.
 XX PA (ADLER/) ADLER D A.
 XX PA (SHEP/) SHEPPARD P O.
 XX PI Adler DA, Sheppard PO;
 XX DR WPI: 2002-642378/69.
 XX DR N-PSDB; AAD45050.
 XX PT Novel secreted salivary polypeptide, zsig63, useful as antimicrobial
 XX agent for treating microbial infection, dental carrier, periodontal
 XX disease, thrush gastrointestinal disease, and for aiding digestion -
 PS Claim 10; Page 28; 33pp; English.
 CC The invention relates to human secreted salivary polypeptide designated
 CC as zsig63 and nucleic acid molecules encoding such polypeptides. zsig63
 CC can be used in detecting agonists and antagonists of its activity, and
 CC is also useful as a host defense polypeptide, immune modulating factor,

CC antipathogenic polypeptide, cell-cell signalling molecule, growth factor,
 CC cytokine, or as secreted extracellular matrix associated proteins with
 CC growth factor hormone activity. It is useful for treating conditions
 CC associated with pathological microbes, including bacterial, fungal and
 CC viral infections, for treating dental caries (tooth decay), periodontal
 CC disease, thrush and gastrointestinal disease, for treating urinary tract
 CC infection, vaginal infection and for preventing infection in skin and
 CC other epithelial wounds, zsig63 is useful for establishing normal
 CC microflora and protect against pathogenic colonisation and invasion, for
 CC treating chronic tissue damage e.g. damage in extremities associated
 CC with diabetes and useful as anti-inflammatory agents. It is useful as a
 CC marker of lung dysfunction, salivary gland dysfunction, or dysfunction of
 CC prostate gland. It is also therapeutically useful for aiding digestion.
 CC Polynucleotides of the invention are used in gene therapy for increasing
 CC or inhibiting zsig63 activity, for detecting abnormalities on human
 CC chromosome 4 associated with disease or other human traits and as
 CC diagnostics in forensic DNA profiling. Sequences of the invention are
 CC useful for stimulating proliferation or differentiation of cardiac
 CC myocytes, for proliferation or differentiation of adipocytes and for
 CC inhibiting chondrosarcomas, atherosclerosis, restenosis and obesity.
 CC The present sequence is human zsig63 protein. zsig63 gene is located
 CC on chromosome 4.

SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 21; Length 219;

Best Local Similarity 100.0%; Pred. No. 2.5e-85;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACIVCAVFAARKRFPIGDDNDGHLPLPSLNIPIGIRNLPPPLYRRPVNTVP 60
 DB 1 MKLLIMACIVCAVFAARKRFPIGDDNDGHLPLPSLNIPIGIRNLPPPLYRRPVNTVP 60
 QY 61 SYPGNTYTDGTGSPYFWITTSQPFVYHIRGFPPLATQNLVPLPRGPFVPSRPFSA 120
 DB 61 SYPGNTYTDGTGSPYFWITTSQPFVYHIRGFPPLATQNLVPLPRGPFVPSRPFSA 120
 QY 121 AAAPAPPIAAEPAPAAPIATATPVAAEPAAAGVAAEPAAEPVAAEPAAEPVAAEPAA 180
 DB 121 AAAPAPPIAAEPAPAAPIATATPVAAEPAAAGVAAEPAAEPVAAEPAAEPVAAEPAA 180
 QY 181 EAPVGEPAAPAEPSPAEPATAKPAAPDEPHSPSLLEQANQ 219
 DB 181 EAPVGEPAAPAEPSPAEPATAKPAAPDEPHSPSLLEQANQ 219

RESULT 3

ABG31608 ID ABG31608 standard; Protein; 219 AA.

AC ABG31608;

DT 15-NOV-2002 (first entry)

DE Human secreted salivary protein zsig63.

KM Human, secreted salivary protein; zsig63; immunogen; zsig63-cytokine;
 KM antibody-cytokine; in vivo killing; pathological microbe; bacteria;
 KM fungal; viral; infection; salivary gland; anti-microbial; dental caries;
 KM tooth decay; periodontal disease; thrush; gastrointestinal disease;
 KM urinary tract infection; vaginal infection; skin infection; microflora;
 KM epithelial wound; pathogenic colonisation; invasion; pro-inflammatory;
 KM chronic tissue damage; vascular system; diabetes; anti-inflammatory;
 KM incompetent immune system; AIDS; acquired immunodeficiency syndrome;
 KM chemotherapy; radiation treatment; lung infection; cystic fibrosis;
 KM digestion; chromosome 4.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..15 /note= "Fusion protein peptide, specifically claimed
 FT in claim 18"

FT Region 1..219 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 9..204 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 14..19 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..21 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..37 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..219 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 17..33 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 24..29 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 24..33 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 25..30 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 38..126 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 66..73 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 103..108 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 109..215 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 124..133 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 127..219 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 134..138 /label= Repeat_2
 FT 139..143 /label= Repeat_3
 FT 144..148 /label= Repeat_4
 FT 149..153 /label= Repeat_5
 FT 154..158 /label= Repeat_6
 FT 159..163 /label= Repeat_7
 FT 164..168 /label= Repeat_8
 FT 169..173 /label= Repeat_9
 FT 174..178 /label= Repeat_10
 FT 179..183 /label= Repeat_11
 FT 184..188 /label= Repeat_12
 FT 187..192 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT 189..193

FT FT /label= Repeat_13
 FT Region 190..197
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 194..198
 FT /label= Repeat_14
 FT Region 199..203
 FT /label= Repeat_15
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 202..215
 FT /label= Repeat_16
 FT /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 204..208
 FT /label= Repeat_16
 FT US2002081701-A1.
 PD 27-JUN-2002.
 XX
 XX
 XX 03-AUG-2001; 2001US-0922480.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX 17-MAR-2000; 2000US-0527345.
 XX
 PA (ADLER/) ADLER D A.
 PA (SHEP/) SHEPPARD P O.
 XX
 PI Adler DA, Sheppard PO;
 XX
 XX WPI; 2002-635468/68.
 DR N-PSDB; ABS52633, ABS52634.
 XX
 PT Novel secreted salivary protein, zsig63 and polynucleotide encoding it
 PT useful for treating microbial infections, inflammatory conditions,
 PT dental caries and lung infections associated with cystic fibrosis
 XX
 PS Claim 10; Page 28; 33pp; English.

CC The present invention relates to a new secreted salivary protein, zsig63.
 CC The invention is useful for detecting, in a test sample, the presence of
 CC an antagonist or agonist of zsig63 protein activity. The invention is
 CC also useful as an immunogen for producing an antibody to zsig63
 CC polypeptide. zsig63-cytokine fusion proteins or antibody-cytokine fusion
 CC protein are useful for enhancing in vivo killing of target tissues.
 CC Pharmaceutical composition comprising purified zsig63 polypeptide are
 CC useful in the treatment of conditions associated with pathological
 CC microbes, including bacterial, fungal and viral infections. High
 CC expression of zsig63 in salivary gland suggests that anti-microbial
 CC polypeptides are useful for treatment of dental caries (tooth decay),
 CC periodontal disease, thrush and gastrointestinal disease. Other
 CC applications can be used in urinary tract infections, vaginal infections,
 CC prevention of infection in skin and other epithelial wounds. The
 CC polypeptides can be used to establish normal microflora and protect
 CC against pathogenic colonisation and invasion. The invention is useful
 CC when pro-inflammatory activity is desired. Applications for
 CC such pro-inflammatory activity include the treatment of chronic tissue
 CC damage, particularly in areas having a limited or damaged vascular system
 CC e.g., damage in extremities associated with diabetes. Antagonists to
 CC zsig63 polypeptides may be useful as anti-inflammatory agents. The
 CC invention is useful for the treatment of patients having incompetent
 CC immune system, such as AIDS (acquired immunodeficiency syndrome) patients
 CC or individuals that have undergone chemotherapy, radiation treatment. The
 CC invention is also useful for the treatment of lung infections associated
 CC with cystic fibrosis and its agonists or antagonists are useful for
 CC aiding digestion. The present amino acid sequence represents the human
 CC secreted salivary protein zsig63 of the invention. This sequence is
 CC encoded by the human zsig63 gene located on chromosome 4.
 XX
 SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLLMACIVCAVAFARKRFRPFIIGDDNDGDGHLPLSPINIPYGINRLPPLTYRRPVNTVP 60
 DB 1 MGLLMACIVCAVAFARKRFRPFIIGDDNDGDGHLPLSPINIPYGINRLPPLTYRRPVNTVP 60
 QY 61 SYRGNTYTDYGLPSYPWILITSPBPYYVYHIRGFPLATOLNVPPLPRGPFYPSRFFSA 120
 DB 61 SYRGNTYTDYGLPSYPWILITSPBPYYVYHIRGFPLATOLNVPPLPRGPFYPSRFFSA 120
 QY 121 AAAPAAPPIAAEPAPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 DB 121 AAAPAAPPIAAEPAPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 QY 181 EAPVGEPAEPEPSPAEPTAKPAEPAPPEPSPLEQANQ 219
 DB 181 EAPVGEPAEPEPSPAEPTAKPAEPAPPEPSPLEQANQ 219

RESULT 4
 AAU74536
 ID AAU74536 strand; Protein; 219 AA.
 XX
 XX AAU74536;
 XX
 XX 23-APR-2002 (first entry)
 XX
 DE Human zsig63 polypeptide.
 XX
 XX Human; zsig63; chromosome 4q12-q13; salivary protein; antimicrobial;
 XX microbial infection; tooth decay; periodontal disease; thrush; emphysema;
 XX gastrointestinal disease; urinary tract infection; vaginal infection;
 XX skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
 XX acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
 XX chronic bronchitis; gene therapy; protein therapy.
 XX
 OS Homo sapiens.
 XX
 XX US6331413-B1.
 XX
 PD 18-DEC-2001.
 XX
 XX 17-MAR-2000; 2000US-0527345.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 PA
 PI Adler DA, Sheppard PO;
 XX
 XX WPI; 2002-096707/13.
 DR N-PSDB; AAS20591.
 XX
 PT Polynucleotides encoding salivary proteins useful as anti-microbial
 PT agents -
 XX
 PS Claim 1; Column 49-52; 29pp; English.

CC The invention relates to a polynucleotide derived from the 4q12-4q13
 CC region of human chromosome 4 and encoding a zsig63 polypeptide, a
 CC secreted salivary protein with anti-microbial activity. Due to their
 CC microbial activity, the sequences can be used in the study of microbial
 CC infections, e.g. for recombinant production of anti-microbial proteins.
 CC The sequences can be used in the treatment of tooth decay, periodontal
 CC disease, thrush, gastrointestinal disease, urinary tract infections,
 CC vaginal infections, skin infections, epithelial wounds, chronic tissue
 CC damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung
 CC infections, sarcoidosis, emphysema and chronic bronchitis. This sequence
 CC represents human zsig63.
 XX
 SQ Sequence 219 AA;

Query Match 100.0%; Score 1178; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DR N-PSDB; ABX93594.

XX Novel isolated zsig63 polypeptide, member of the adhesin family, useful
PT for treating dental caries, periodontal disease, thrush,
PT gastrointestinal disease, urinary tract infections, vaginal infections,
PT skin infections

XX Claim 10; Page 27-28; 32pp; English.

XX The invention relates to an isolated zsig63 polypeptide comprising at
CC least 90% identity to an amino acid sequence which comprises domain 1 of
CC zsig63, domain 2, domain 3, mature zsig63 and full length zsig63. Also
CC included are the polynucleotide encoding zsig63, a zsig63 expression
CC vector, a cultured cell comprising the vector and expressing the protein,
CC a DNA encoding a fusion protein (comprising amino acids 1-15, 16-37,
CC 38-126, 127-219 or 16-219 of zsig63 and an additional protein), using a
CC zsig63 reporter gene construct to identify zsig63 agonists, and
CC producing an anti-zsig63 antibody using zsig63 immunogenic peptides.
CC Zsig63 is useful for detecting in a test sample, the presence of
CC antagonist of zsig63 protein activity. Zsig63 has antimicrobial
CC activity and since exhibits high expression in salivary gland, can be
CC used for treating dental caries, periodontal disease, thrush, and
CC gastrointestinal disease, urinary tract infections, vaginal infections,
CC skin infections and other epithelial wounds. The polypeptides can be
CC used to establish normal microflora and protect against pathogenic
CC colonization and invasion. Zsig63 can also be used for providing
CC pro-inflammatory activity for treating chronic, tissue damage
CC particularly in areas having limited or damaged vascular system, e.g.
CC in diabetes, and for treating immunocompromised AIDS patients or in
CC individuals that have undergone chemotherapy, radiation treatment, for
CC treating lung infections e.g. in cystic fibrosis. Detection of zsig63
CC polypeptide at relatively high levels in the trachea may indicate that
CC such polypeptides may serve as a marker of lung dysfunction. Zsig63 is
CC also useful in diagnosing conditions associated with salivary gland or
CC lung dysfunction including salivary gland carcinoma, Pneumocystis carinii
CC infection, emphysema, chronic bronchitis, prostate dysfunctions such
CC as prostatic adenocarcinoma, aiding digestion, and as components of
CC defined cell culture media and may be used to replace serum that is
CC commonly used in culture. The DNA is useful in gene therapy applications
CC to increase or inhibit zsig63 activity, and for detecting abnormalities
CC on human chromosome 4 (e.g. 4q12-4q13, associated with dentinogenesis
CC imperfecta, and dentin dysplasia type II). Zsig63 is an adhesin family
CC member. The present sequence represents human zsig63.

XX Sequence 219 AA;

SQ Query Match 100.0%; Score 1178; DB 24; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.5e-85;

Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLLMACTVCVAFARRRPPIGDDNDGHPHPSLNIPIGIRNLPPPLYRRPVNTVP 60
DB 1 MLLLMACTVCVAFARRRPPIGDDNDGHPHPSLNIPIGIRNLPPPLYRRPVNTVP 60
QY 61 SYRGNTYTDGTPSYWILTSRGFPYVYHIRGFPLATQNLVPPPLPRGFPFVPSRFSFA 120
DB 61 SYRGNTYTDGTPSYWILTSRGFPYVYHIRGFPLATQNLVPPPLPRGFPFVPSRFSFA 120
QY 121 AAAPAPPIAAEPAAAAPLTATPVAAPAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180
DB 121 AAAPAPPIAAEPAAAAPLTATPVAAPAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180
QY 181 EAPVGEPAAEPSPAEPATAKPAADPHPSLSLEQANQ 219
DB 181 EAPVGEPAAEPSPAEPATAKPAADPHPSLSLEQANQ 219

RESULT 6
AAW30653
ID AAW30653 standard; Protein; 219 AA.
XX
AC AAW30653;
XX

DT 12-APR-1999 (first entry)

XX Human secreted protein clone cp16 1 protein.
XX
XX Human; secreted protein; nutritional activity; cytokine; vaccine;
KW cell proliferation; differentiation; immune stimulation; suppression;
KW haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
KW chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
KW tumour invasion suppression; tumour inhibition.

XX Homo sapiens.

XX MO9901466-A1.

XX 14-JAN-1999.

XX 01-JUL-1998; 98MO-US13813.

XX 27-OCT-1997; 97US-0958304.

XX 02-JUL-1997; 97US-0887195.

XX (GENEX) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;

PI Spaulding V, Treacy M;

DR WPI; 1999-105994/09.

DR N-PSDB; AAW80740.

PT New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult testes, adult brain, foetal kidney,
PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
PT potential vaccines

PS Claim 24; Page 71-72; 107pp; English.

XX The present sequence represents a human secreted protein from clone
CC cp16 1, deposited as ATCC 98482. Human secreted protein clone
CC polynucleotides and proteins are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Suggested
CC activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccine) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy.

XX Sequence 219 AA;

SQ Query Match 99.5%; Score 1172; DB 20; Length 219;
Best Local Similarity 99.1%; Pred. No. 7.4e-85;

Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLLMACTVCVAFARRRPPIGDDNDGHPHPSLNIPIGIRNLPPPLYRRPVNTVP 60
DB 1 MLLLMACTVCVAFARRRPPIGDDNDGHPHPSLNIPIGIRNLPPPLYRRPVNTVP 60
QY 61 SYRGNTYTDGTPSYWILTSRGFPYVYHIRGFPLATQNLVPPPLPRGFPFVPSRFSFA 120
DB 61 SYRGNTYTDGTPSYWILTSRGFPYVYHIRGFPLATQNLVPPPLPRGFPFVPSRFSFA 120
QY 121 AAAPAPPIAAEPAAAAPLTATPVAAPAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180
DB 121 AAAPAPPIAAEPAAAAPLTATPVAAPAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180
QY 181 EAPVGEPAAEPSPAEPATAKPAADPHPSLSLEQANQ 219
DB 181 EAPVGEPAAEPSPAEPATAKPAADPHPSLSLEQANQ 219

RESULT 7
AAV19472
ID AAV19472 standard; Protein; 221 AA.
XX
AC AAV19472;
XX
DT 14-JUL-1999 (first entry)
XX
DE Amino acid sequence of a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; chymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS.
XX
OS Homo sapiens.
XX
FN WO9922243-A1.
XX
PD 06-MAY-1999.
XX
PE 23-OCT-1998; 98MO-US22376.
XX
PR 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.
PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063097.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Peng P, Florence C, Florence KA, Greene JM, Janat P;
PI Kaye H, Lafleur DM, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
DR WPI; 1999-303069/25.
DR N-PSDB; AAX61352.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX
PS Claim 11; Page 401-402; 546pp; English.
XX
XX The specification describes cDNA sequences (AAV61322-X61470) encoding
XX human secreted proteins (AAV19472-Y19590). The polynucleotides and their
XX corresponding secreted polypeptides are useful for preventing, treating
XX or ameliorating medical conditions, e.g. by protein or gene therapy.
XX Pathological conditions can also be diagnosed by determining the amount
XX of the polypeptides in a sample or by determining the presence of
XX mutations in the polynucleotides. Specific uses are described for each
XX of the polynucleotides, based on which tissues they are most highly
XX expressed in, and include developing products for the diagnosis or
XX treatment of cancer, tumours, neurodegenerative disorders, developmental
XX abnormalities and fetal deficiencies, blood disorders, leukemia,
XX diseases of the immune system, autoimmune diseases, hepatic and renal
XX disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
XX and cognitive disorders, schizophrenia, prostate diseases, obesity,

CC disorders involving osteoclasts such as osteoporosis, arthritis or
CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
CC disorders, infections and AIDS. The polypeptides are also useful for
CC identifying their binding partners.
XX
XX SQ Sequence 221 AA;
XX
XX Query Match 94.5%; Score 1113.5; DB 20; Length 221;
XX Best Local Similarity 95.0%; Pred. No. 3.1e-80;
XX Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;
XX
QY 1 MKLLIMACIVCAFAKRRPFPIGEDNDGDHPLPSINIPYGINLPPLLYRPVNTVP 60
DB 1 MKLLIMACIVCAFAKRRPFPIGEDNDGDHPLPSINIPYGINLPPLLYRPVNTVP 60
QY 61 SYPGNTYDTGTPSYPMILTSFGPPYVYHNGFPLATOLANPPLPRGPPFPSPRFSFA 120
DB 61 SYPGNTYDTGTPSYPMILTSFGPPYVYHNGFPLATOLANPPLPRGPPFPSPRFSFA 120
QY 121 AAAPAPPIAEPAAAPLTAIPVAEPAGAPVAEPAAEP- VCAEPAAEPVAAEPA 179
DB 121 AAAPAPPIAEPAAAPLTAIPVAEPAGAPVAEPAGXGCHLELEPAEAPVAAEPA 180
QY 180 AEAPEVGEPAABEPPSPAPATAPKPAEPHPSPSLEQANQ 219
DB 181 AEAPEVGEPAABEPPSPAPATAPKPAEPHPSPSLEQANQ 220
XX
XX RESULT 8
XX ABG21919
XX ID ABG21919 standard; Protein; 325 AA.
XX
XX AC ABG21919;
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #21910.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS86106.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX PS Claim 20; SEQ ID No 52278; 1033p; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostic, forensic, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 325 AA;

Query Match 15.6%; Score 183.5; DB 22; Length 325;
 Best Local Similarity 34.3%; Pred. No. 9.5e-07;
 Matches 57; Conservative 8; Mismatches 78; Indels 23; Gaps 6;

OY 48 PPPLYYRPVNTVPSYEGNTYTDGLPSYWMILTSQGFYVYHIRGFLATQINVPPLPR 107

DB 45 PCPSRYCPQMTPEPP-----PAPPALPPPPPP-----APASPPLPAPSP- 87

OY 108 GPPFPVPSRFPFAAAPPPIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPVGA 167

DB 88 --PHAPPQALPPPPALPPPP--ASPPPLPPALSPPLPAPPPSAPASPPAPPPPPAP 143

OY 166 PAEAAPVAAEPAAEPVGAEPAAEPSPAPAPAPAPAPAPAPAPAPAPAPAPAP 212

DB 144 PPSPPAPPPSAPSPAPLPAPAPSPSPAPPPAPPPAPPPAPPPAPPPAPPPAPPP 168

RESULT 9

AA040797
 ID AA040797 standard; Protein; 207 AA.

XX AA040797;

XX 25-MAR-2003 (updated)

DT 08-FEB-1994 (first entry)

XX Sequence of a 10-amino acid repetitive sequence of an antigenic

DE peptide domain encoded by an insert of clone Tcd.

XX Epitope repeat; antigen; Chagas disease; TCD.

XX Trypanosoma cruzi.

OS Key Location/Qualifiers

FT MISC-difference 42 /label= Degeneracy in repeat unit

FT /note= "see also AAs 88,108,158,162"

XX WO9316199-A1.

XX 19-AUG-1993.

XX 11-FEB-1993; 93WO-US01231.

XX 14-FEB-1992; 92US-0836642.

XX (REBD/) REBD S G.

XX Reed SG;

XX WPI; 1993-272900/34.

XX N-PSDB; AA040797.

XX Antigenic peptide domain of trypanosoma cruzi - used to diagnose

XX or immunise against infection, and screen blood supplies

PS Disclosure; Figure 1; 16pp; English.

XX Clone Tcd encodes a 10-amino acid repetitive sequence present
 CC in 20.5 copies with minor degeneracies present in 5 posns.
 CC The predicted mol. wt. of recombinant unglycosylated Tcd antigen
 CC is 36.3kD. Analysis of DNA from seven geographically diverse
 CC T. cruzi isolates indicate that Tcd gene sequence was conserved.
 CC A 636 bp fragment of clone Tcd was used to probe DNA from several
 CC other protozoan parasites of humans. However, it did not hybridise
 CC to T. brucei, Leishmania chagasi, L. amazonensis, L. donovani
 CC or T. rangeli.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 207 AA;

Query Match 15.3%; Score 180; DB 14; Length 207;
 Best Local Similarity 38.4%; Pred. No. 1.1e-06;
 Matches 43; Conservative 11; Mismatches 46; Indels 12; Gaps 2;

OY 114 PSRFSMAAAPAP-----PIAAEPAAAPLTATVAAEPAAAGAVAAEPAAEPV 164

DB 85 PAESKSAEPKPAEPKSAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPK 144

OY 165 GAEPAAEPVAAEPAAEPVGAEPAAEPSPAPAPAPAPAPAPAPAPAPAPAPAPAP 213

DB 145 PAEPKSAEPKPAEPKSAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPKPAEPK 196

RESULT 10

AA060915
 ID AA060915 standard; Protein; 207 AA.

XX AA060915;

XX 02-APR-1997 (first entry)

XX T. cruzi Tcd antigen.

XX Tcd; diagnosis; infection; antigen; Chagas disease.

XX Trypanosoma cruzi.

OS Key Location/Qualifiers

FT Region 6..20 /label= Antigenic epitope

FT /note= "Claim 3, page 38"

XX WO9629605-A2.

XX 26-SEP-1996.

XX 12-MAR-1996; 96WO-US03380.

XX 14-MAR-1995; 95US-0403379.

XX (COR-) CORIXA CORP.

XX Reed SG;

XX WPI; 1996-485445/48.

XX N-PSDB; AAT46149.

XX Detecting and preventing T. cruzi infection - using polypeptide(s)

XX or antibodies contg. or reactive with antigen epitope(s) of T. cruzi

XX proteins

XX Disclosure; Fig 4; 59pp; English.

XX The Tcd antigen (AA060915) of Trypanosoma cruzi includes an

XX antigenic epitope (see also AA060917) that can be utilised, pref.

XX with other T. cruzi epitopes (see also AA060914-19), in immunosays

XX for T. cruzi infection (Chagas' disease) and in vaccine compns.

CC shock syndrome), inflammatory conditions (e.g., measles, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 903 AA;

Query Match 14.9%; Score 175.5; DB 23; Length 903;
 Best Local Similarity 32.3%; Pred. No. 1.2e-05;
 Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;

QY 35 HPSLNIPIGIRNLP---PPLYRRPVNTVPSYSGNTYTDGLPSYWILTSRGPVYHIR 91
 DB 579 HKQVNVPMNTSSPTSNPVTTTKPVTTT-----KVVTTTKP-----VTTTKP----- 622
 QY 92 GFPLATQLVNPPPLPRGPFVPSRFFSAAAPA-APPIAEPAAAPVLTATPVAEPAA 150
 DB 623 ----VTIINQPSVVP-----AAKPAAPAKPVAAKPVATKATVAP----- 658
 QY 151 GAPVAEPAAEA-PVGAEPAAEPVAAEPAAEPVGAEPAAEPBAPATKPAAPBPH 209
 DB 659 --PVAVKATATAKPVAAKPAAVRPPAA--AAAKPVATKEVVRPQAAKPAATKPAATKPM 714
 QY 210 PPSPLE 215
 DB 715 VKMSRE 720

RESULT 15

AA040303 ID AAM40303 standard; Protein; 3177 AA.

AC AAM40303;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3448.

XX Human, nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoclastic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0596042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;

XX Zhao QX, Zhou P, Goodrich R, Drmanac RT;

XX MPI; 2001-442253/47.

XX N-PSDB; AA159459.

XX Example 6; SEQ ID NO 3448; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 3177 AA;

Query Match 14.9%; Score 175.5; DB 22; Length 3177;
 Best Local Similarity 32.3%; Pred. No. 4.4e-05;
 Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;

QY 35 HPSLNIPIGIRNLP---PPLYRRPVNTVPSYSGNTYTDGLPSYWILTSRGPVYHIR 91

DB 2853 HKQVNVPMNTSSPTSNPVTTTKPVTTT-----KVVTTTKP-----VTTTKP----- 2896

QY 92 GFPLATQLVNPPPLPRGPFVPSRFFSAAAPA-APPIAEPAAAPVLTATPVAEPAA 150

DB 2897 ----VTIINQPSVVP-----AAKPAAPAKPVAAKPVATKATVAP----- 2932

QY 151 GAPVAEPAAEA-PVGAEPAAEPVAAEPAAEPVGAEPAAEPBAPATKPAAPBPH 209

DB 2933 --PVAVKATATAKPVAAKPAAVRPPAA--AAAKPVATKEVVRPQAAKPAATKPAATKPM 2988

QY 210 PPSPLE 215

DB 2989 VKMSRE 2994

Search completed: January 15, 2004, 15:13:23
 Job time : 43 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:10:34 : Search time 35 Seconds
(without alignments)
1614.672 Million cell updates/sec

Title: US-09-923-236-2
Perfect score: 1178
Sequence: 1 MKLLMNCIVVAFARRRF.....TAKPAAPHPHSPSLQANQ 219

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL.23.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_tvirus:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.5	30.5	344	11 Q9D5J3	Q9D5J3 mus musculus
2	224.5	19.1	3179	12 Q8V2A4	Q8V2A4 human herpes
3	221	18.8	1100	2 Q9X6J9	Q9X6J9 klebsiella
4	214.5	18.2	1480	10 Q9LIR8	Q9LIR8 arabidopsis
5	214	18.2	3105	12 Q8U2T9	Q8U2T9 cercopithec
6	206.5	17.5	607	5 Q17112	Q17112 babesia bov
7	204.5	17.4	1070	2 Q8GFS8	Q8GFS8 citrobacter
8	199	16.9	391	5 Q268J9	Q268J9 trypanosoma
9	194.5	16.5	722	5 Q268J9	Q268J9 trypanosoma
10	193	16.4	3534	12 Q39266	Q39266 equine herpes
11	189.5	16.0	1108	2 Q49542	Q49542 mycoplasma
12	188	16.0	839	16 Q9RX57	Q9RX57 deinococcus
13	186.5	15.8	395	16 Q9RV74	Q9RV74 deinococcus
14	186.5	15.8	616	11 Q99KJ1	Q99KJ1 mus musculus
15	183	15.5	744	10 Q65375	Q65375 arabidopsis

17	182	15.4	386	10 Q9FP05	Q9FP05 chlamydomon
18	181	15.4	928	12 Q9IKX9	Q9IKX9 cercopithec
19	181	15.4	2657	11 Q88493	Q88493 mus musculus
20	178.5	15.2	243	10 Q9X1V1	Q9X1V1 cucumis sat
21	177.5	15.1	598	16 Q8VKN7	Q8VKN7 mycobacteri
22	177	15.0	1315	10 Q9SPM0	Q9SPM0 zea mays (m
23	176.5	15.0	3084	12 Q8U2T1	Q8U2T1 pseudorabies
24	176	14.9	959	16 Q8XRH0	Q8XRH0 ralsconia s
25	175	14.9	857	16 Q981K2	Q981K2 rhizobium l
26	174.5	14.8	979	4 Q8N4Z1	Q8N4Z1 homo sapien
27	174.5	14.8	1611	3 Q42854	Q42854 schizosacch
28	174	14.8	236	10 Q9LM00	Q9LM00 pinus taeda
29	174	14.8	503	12 Q39779	Q39779 equine herp
30	173.5	14.7	786	10 Q48809	Q48809 arabidopsis
31	173	14.7	1174	4 Q94854	Q94854 homo sapien
32	171.5	14.6	438	16 Q9A2M5	Q9A2M5 caulobacter
33	171.5	14.6	616	4 Q9H6K5	Q9H6K5 homo sapien
34	171	14.5	544	12 Q89392	Q89392 parametium
35	171	14.5	1340	16 Q9LJH8	Q9LJH8 streptomyce
36	170.5	14.5	601	5 Q17113	Q17113 babesia bov
37	170	14.4	295	2 Q32850	Q32850 mycobacteri
38	170	14.4	428	10 Q941Q8	Q941Q8 zea mays (m
39	170	14.4	548	16 Q06404	Q06404 mycobacteri
40	170	14.4	842	5 Q9VG59	Q9VG59 drosophila
41	170	14.4	864	5 Q95U45	Q95U45 drosophila
42	170	14.4	864	5 Q9VGC8	Q9VGC8 drosophila
43	170	14.4	1354	11 Q9EPW8	Q9EPW8 mus musculu
44	169.5	14.4	802	11 P70433	P70433 mus musculu
45	169	14.3	312	5 Q9UAN1	Q9UAN1 drosophila

ALIGNMENTS

RESULT 1
ID Q9D5J3 PRELIMINARY; PRT; 344 AA.
AC Q9D5J3.
DT 01-JUN-2001 (TRMBLrel. 17, Created)
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRMBLrel. 17, Last annotation update)
DE 4930432K09R1K protein.
GN 4930432K09R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukumitsu Y., Kono H., Adachi U., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Fujino M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gietzung S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombert P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmig L.,
RA Wyshew-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohlsch S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:665-690(2001).
DR EMBL; AK015291; BAB29782.1; -
DR MGI; MGI:1921029; 4930432K09R1K.

DR InterPro: IPR002965; P rich_extensn.

DR PRINTS; PR01217; PRICHEXTENS. D65EAD71CE5802AC CRC64;

Query Match
Best Local Similarity 32.0%; Score 359.5; DB 11; Length 344;
Matches 106; Conservative 25; Mismatches 87; Indels 113; Gaps 11;

QY 1 MKLLMAGTCAVAFARRRPPPTIGEDD--DDGHPLHPSLNIPIYGI--RNLV---PPLY 53
DB 1 MKLLMAGTCAVAFARRRPPPTIGEDD--DDGHPLHPSLNIPIYGI--RNLV---PPLY 59
QY 54 RPVNTVPSYGNVYTGTLPSYPMILTSFGFVYHIRGFLATOLNVPPL----- 104
DB 60 QCGNTTKTTPGSELEKGLSLYPMIATPSKLTIVYOSLNPADAPLNGPPVAPLPKAF 119
QY 105 PRGPFVPPSRRF-----SAAAPAAAP----- 127
DB 120 PRGPFVPPSRRF-----SAAAPAAAP----- 127
QY 128 -----DIAEPA-----AAPLTATVAAEPAAGAPVAAE 157
DB 180 PSGRPLADEPALSPGAPAPVQCGAPAPVHQLAAVDPASRLMADEPAVPLSVGAQ 239
QY 158 PAEAPVVG-----APPAAPVPAEPAAPV-----GVEPAEESP----- 195
DB 240 SLAESDAGISPENKPTSGSPAATQS.PALPAAGLAIVEAKLPAESAGRPPEALMASOS 299
QY 196 -----APPATKPAAPPPSPSLQANQ 219
DB 300 VVGKLTAEPTAKPOVLEVEAKSASQEAQ 330

RESULT 2

Q8V2A4 PRELIMINARY; PRT; 3179 AA.

AC Q8V2A4; PRELIMINARY; PRT; 3179 AA.
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE BPLF.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M-ABA;
RT "Schmaus S., Wolf H., Schwarzmann F.;
RT "The open reading frame BPLF of Epstein-Barr virus is expressed in
RT late viral replication."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF448220; AAL40834.1; -
DR InterPro: IPR006928; Herpes teg N.
DR InterPro: IPR002965; P rich_extensn.
DR Pfam: PF04843; Herpes_teg_N; 1.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 3179 AA; 340397 MW; AB0A5D9088A945B1 CRC64;

Query Match
Best Local Similarity 31.1%; Score 224.5; DB 12; Length 3179;
Matches 68; Conservative 23; Mismatches 61; Indels 67; Gaps 9;
QY 24 GEDNDGDGHLPLSLNIPYGINLPPLYRPNVNTVPSYGNVYTGTLPSYPMILTSFG 83
DB 278 GEAAGTGGADSSPPWLPFRRIKXPNL--RPL-----PSRFTS---DSFPAARYSP- 325
QY 84 FPYVYHIRGFLATOLNVPPLPRGPFVPPSRRFSAAPAAAPVPAEPAAPVPAEPA 142
DB 326 -----AKTNSPPSSP-----ASAPASAPASAPASAPASAPASAPASAP 359
QY 143 PVAEPAAGPVAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 193

DB 360 PASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAP 419

QY 194 -SPAEPATKRP-----AAPF-----PRPS 213
DB 420 HTPGVAPSTPSPRAGGAPQTPKRRKGLGKDSPKKPT 458

RESULT 3

Q9X6J9 PRELIMINARY; PRT; 1100 AA.

AC Q9X6J9; PRELIMINARY; PRT; 1100 AA.
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE DNA primase.
GN PR1.
OS Klebsiella oxytoca.
OC Plasmid pACM1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334746; PubMed=10873523;
RA Preston K.B., Radomski C.C.A., Venezia R.A.;
RT "Nucleotide sequence of a 7-kb fragment of pACM1 encoding an IncM DNA
RT primase and other putative proteins associated with conjugation."
RL Plasmid 44:12-23(2000).
DR EMBL; AF139719; AAD31802.1; -
KW Plasmid.
SQ SEQUENCE 1100 AA; 120370 MW; 36EAD2BD33FCB732 CRC64;

Query Match
Best Local Similarity 31.4%; Score 221; DB 2; Length 1100;
Matches 76; Conservative 19; Mismatches 85; Indels 62; Gaps 9;

QY 9 IYCVAFARRRRP-----FIGEDD--NND-----GHPLHPSLNIPIYGIN 46
DB 296 MVEVAGRLKDTFPNSTHFLADNDYKDENVGLKATAEALTAGHVLVPAASN----- 352
QY 47 LPPLLYRPVNTVPSYGNVYTGTLPSYPMILTSFGFVYHIRGFLATOLNVPPLP 105
DB 353 -----KEGLTDYNDLHVEGLEQVLAQVEG--AINQNRVPTM 388
QY 106 PRGPFVPPSRRFSA-AAAPAP-----PLAEPAAAPVLTATPVAAEPAAGPVAEPA 160.
DB 389 PTDNPNITDVNHSSTDSAAVAAPBEKAAPVASTPAAPVEEAPVASAPAAEVAEAPVA 448
QY 161 EAPVGAE-----PAEAPVAAEPAEAPVGVPEPAEESPAPAPATKRPAAEPHPSPL 215
DB 449 SAPAAELVETAPVASTPEAAEPVAAEPVAAEPVAAEPVAAEPVAAEPVAAEPVAA 508
QY 216 QA 217
DB 509 SA 510
RESULT 4
Q9LIE8 PRELIMINARY; PRT; 1480 AA.
AC Q9LIE8; PRELIMINARY; PRT; 1480 AA.
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Similarity to cell wall-plasma membrane linker protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;

RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=2036309; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
Sequence features of the regions of 4,251,635 bp covered by ninety pl,
TAC and BAC clones.";
RL DNA Ref. 7:217-221(2000).
DR EMBL; AP001306; BAB03062.1; -
DR HSSP; P24337; IAHY.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF00234; try_p_alpha_amy1.1.
DR PRINTS; P01217; PRICEXTENS.
SQ SEQUENCE 1480 AA; 147154 MW; D1AC0C79F155E732 CRC64;

Query Match 18.2%; Score 214.5; DB 10; Length 1480;
Best Local Similarity 32.6%; Pred. No. 4,8e-07; Indels 47; Gaps 9;
Matches 63; Conservative 13; Mismatches 70;

QY 33 PLHSLNIPYGRNLP---PPLYRPNVTVPSPGNTYDTGLPSYPMILSPGFPYVYH 89
DB 665 PAKPVPVAP-----PIATPPVAKPVPATPP-----TATPP-----IATP----- 698
QY 90 IRGFPLATQLVNPPV--PPRGFPV-----PSGRFSAAPAAAP--PPIAEPAAAPL 139
DB 699 ----PIAT-----PVPVTPPTATPPVATPPPIAKPPTTPTATPPVAMPPIATPPPAKPP 750
QY 140 TATPPAAPPAAGAPPAAPPAAPVAGAPPAAPPAAPPAAPVAGVPPAAEPPAPPA 199
DB 751 ATPPIATPPVAKPVPVTPPTATPPPIAKSPVATPTATPPVATPPVATPPVATPPPT 810
QY 200 TAKPAAPPPPS 212
DB 811 TAPPTATPPVAKP 823

RESULT 5
Q8UZ19 PRELIMINARY; PRT; 3105 AA.
ID Q8UZ19
AC Q8UZ19
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE BLPI1.
OS Cercopithecine herpesvirus 15.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=104228;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=97048062; PubMed=8892903;
RA Franken M., Devergne O., Rosenzweig M., Annis B., Kieff E., Wang F.;
RT "Comparative analysis identifies conserved tumor necrosis factor
receptor-associated factor 3 binding sites in the human and simian
Epstein-Barr virus oncogene LMP1.";
RT Epstein-Barr virus oncogene LMP1.";
RL J. Virol. 70:7819-7826(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=99412410; PubMed=10482645;
RA Rivallier P., Quink C., Wang F.;
RT "Strong selective pressure for evolution of an Epstein-Barr virus
LMP2B homologue in the rheus lymphocryptovirus.";
RT J. Virol. 73:8867-8872(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=20304984; PubMed=10846073;

RA Jiang H., Cho Y.-G., Wang F.;
RT "Structural, functional, and genetic comparisons of Epstein-Barr virus
nuclear antigen 3A, 3B, and 3C homologues encoded by the rheus
lymphocryptovirus.";
RL J. Virol. 74:5921-5932(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=20440633; PubMed=10970361;
RA Rao P., Jiang H., Wang F.;
RT "Cloning of the rheus lymphocryptovirus viral capsid antigen and
Epstein-Barr virus-encoded small RNA homologues and use in diagnosis
of acute and persistent infections.";
RL J. Clin. Microbiol. 38:3219-3225(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RX MEDLINE=21602573; PubMed=11739708;
RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;
RT "Complete Nucleotide Sequence of the Rheus Lymphocryptovirus: Genetic
Validation for an Epstein-Barr Virus Animal Model.";
RL J. Virol. 76:421-426(2002).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Moghaddam A., Koch J., Annis B., Wang F.;
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Moghaddam A., Annis B., Wang F.;
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Rivallier P., Quink C., Wang F.;
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Jiang H., Wang F.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Rao P.V., Jiang H., Wang F.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RX EMBL; AY037858; AAK95420.1; -
DR InterPro; IPR005613; AIP3.
DR InterPro; IPR006928; Herpes teg N.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF03915; AIP3; 1.
DR Pfam; PF04843; Herpes teg N; 1.
DR PRINTS; P01217; PRICEXTENS.
SQ SEQUENCE 3105 AA; 332890 MW; 7C6AE4D7D99FC7C3 CRC64;

Query Match 18.2%; Score 214; DB 12; Length 3105;
Best Local Similarity 29.6%; Pred. No. 1.1e-06;
Matches 61; Conservative 18; Mismatches 105; Indels 22; Gaps 7;

QY 18 RRPFGDDNDGHPHPSINIPYGRNLPPLVY--RPVNTVPSPGNTYDTGLPSY 75
DB 390 RRLMSSTTDEDQDPP--PPRRPSPVLRLLAPVTVVQGPAPVPTQASVPPSAAPPH 447
QY 76 PWILTSPGFPVYVHIR---GFPLATQLVNPPVLPDRGFPVPPSRRFSAAPAAAP--PI 129
DB 448 P-----SPVIRPHSPTRKPPDPPTTSQLPPTQPPPISTPLVPP-----ISTPLVPRPT 497

QY 130 AEPAAAPLTAATPVAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAAAPVGEPA 189
 DB 498 POPPAAPPTPOPPOAATPTPOPPOAATPTPOPPOAATPTPOPPOAATPTPOPPOAATPT 557
 QY 190 AEPSPAE--PATAKPAEPHSPS 213
 DB 558 POPPOATSHAPOLPRAASAPPOPTPT 583

RESULT 6

Q17112 PRELIMINARY; PRT; 607 AA.
 ID 017112:
 AC 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE 80 kDa protein.
 OS Babesia bovis.
 OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
 OX NCBI_TaxID=5865;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93341567; PubMed=8341317;
 RA Dalrymple B.P., Peters J.M., Goodger B.V., Bushell G.R.,
 RA Waltsbuhl D.J., Wright I.G.,
 RT "Cloning and characterisation of cDNA clones encoding two Babesia
 RT bovis proteins with homologous amino- and carboxy-terminal domains.";
 RT Mol. Biochem. Parasitol. 59:181-189 (1993).
 RL EMBL; M93126; AAA02753.1; -
 SQ SEQUENCE 607 AA; 67129 MW; 2E06ECTDFB43D732 CRC64;

Query Match 17.5%; Score 206.5; DB 5; Length 607;
 Best Local Similarity 46.5%; Pred. No. 7.3e-07;
 Matches 46; Conservative 7; Mismatches 35; Indels 11; Gaps 2;

QY 128 PIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAAAPVGE 187
 DB 329 PVEEPVAAEPPIVEEPIVEEPIVEEPIVEEPIVEEPIVEEPIVEEPIVEEPIVEE 388
 QY 188 PAEEP-----SPAEPATAKPA---APEPHSPSLE 215
 DB 389 PVAEPEVAAEPVCEPTPAETPAEKPAETPAEKPAETPAEE 427

RESULT 7

Q8GFS8 PRELIMINARY; PRT; 1070 AA.
 ID 08GFS8:
 AC 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE DNA primase.
 GN PRI.
 OS Citrobacter freundii.
 OC Plasmid pCTX-M3.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Citrobacter.
 OX NCBI_TaxID=546;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Goldbiewski M., Zienkiewicz M., Adamczyk M., Kern-Zdanowicz I.,
 RA Ceglowski F.,
 RT "Complete nucleotide sequence of highly transmissible plasmid pCTX-M3.";
 RT Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF550415; AAN87680.1; -
 KW Plasmid.
 SQ SEQUENCE 1070 AA; 117583 MW; F16IDCB3137B6500 CRC64;

Query Match 17.4%; Score 204.5; DB 2; Length 1070;
 Best Local Similarity 30.9%; Pred. No. 1.7e-06;
 Matches 69; Conservative 16; Mismatches 81; Indels 57; Gaps 7;

QY 9 IYCVAFARKRRP-----FIGBD--ND-----GHPLPSLNIPIYGIN 46
 DB 298 MEVAVGRKLDTFPNSHTYFLDNDYKDENVGLEKATAEITAGHVLVPAFSN----- 352
 QY 47 LPPLLYRPVNTVSPGNTYTDGLPSYPILTSPPPYV-YHIRGFPLATQLVNPLP 105
 DB 353 -----KEGLTDYNLDHVSSEGLEQVRLQVEG--AINQMRVDTM 388
 QY 106 PRGPPVPPSPRFFSAAAPPAAPPIAAAPPLTATVAAEPAAAGAPVAAEPAAAPVGE 165
 DB 389 PTDNENINDVNHSSSTDSA-----AVAAPEKAPVASTPAAEVPAAPVAAAPAAEAVE 443
 QY 166 AEPAAEPVAAEPAAEPVGEPAEPAAEPSPAPATAPKPAEP 208
 DB 444 AAVASAPAAEPVETAVASTPEAAEPVEAAAPVASAPAAEP 486

RESULT 8

Q96579 PRELIMINARY; PRT; 796 AA.
 ID 096579:
 AC 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Surface antigen PHG57#5 (Fragment).
 OS Trypanosoma cruzi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tuiahuen;
 RA Garcia G.A., Bontempi E., Bua J., Ruiz A.M.,
 RT "Molecular characterization of a Trypanosoma cruzi clone recognized by
 RT an anti-Schistosoma japonicum glutathione-S-transferase serum.";
 RT Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF091620; AAC98688.2; -
 FT NON TER 1
 SQ SEQUENCE 796 AA; 83983 MW; DD4DCC06622533D8 CRC64;

Query Match 17.3%; Score 204; DB 5; Length 796;
 Best Local Similarity 30.5%; Pred. No. 1.4e-06;
 Matches 62; Conservative 15; Mismatches 62; Indels 64; Gaps 7;

QY 47 LPPLLYRPV-----NTVPSYP-----GNTYTDGLPSYPILTSP 82
 DB 411 IPPP-ERRPVPAATSSSVBEPANERTVTNTQPTVPSPATAGPQTD----- 456
 QY 83 GPPVYHIRGFPLATQLVNPLPGRGPFV-----PSRFSAAAPPAPIAEP 133
 DB 457 -----OTTVAASVPSGAGSKPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEP 502
 QY 134 AAAAPLTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAEPAAAPVGEPAEP 193
 DB 503 KPABPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 562
 QY 194 SPAEPATAKPAEP---HSPS 213
 DB 563 KPABPKSAEPKPAEPKSAEPKPT 585

RESULT 9

Q26892 PRELIMINARY; PRT; 391 AA.
 ID 026892:
 AC 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Surface antigen (Fragment).
 OS Trypanosoma cruzi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5693;
 RN [1]
 RP SEQUENCE FROM N.A.

RC	STRAIN=CAI;
RX	MEDLINE=92389985; PubMed=1518558;
RA	Buschlazzo A., Campeella O.E.;
RT	"Sequence of the gene for a Trypanosoma cruzi protein antigenic during
RT	the chronic phase of human Chagas disease.";
RL	Mol. Biochem. Parasitol. 54:125-128(1992).
DR	EMBL; M92046; AAB97873.1; -.
FT	NON TER
SQ	SEQUENCE 391 AA; 39459 MW; F234844B1F865BC7 CRC64;
Query Match	16.9%; Score 199; DB 5; Length 391;
Best Local Similarity	43.8%; Pred. No. 1.6e-06;
Matches	42; Conservative 14; Mismatches 40; Indels 0; Gaps 0
Dy	113 PPSRFFSAAAAPPAEPIAEPAAAPLTATPVAAEPPAGAPVAAEPAAEPVGAEPAAEA 172
Db	5 PKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSGEPKSAEPKSAEPKSAEPKAE 64
Dy	173 PVAAEPAAEPVGVGPAAEEPPAAPATTKPAPEP 208
Db	65 PKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 100
RESULT 10	
ID	Q26893 PRELIMINARY; PRT; 722 AA.
AC	Q26893;
DT	01-NOV-1996 (TREMBlrel. 01, Created)
DT	01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT	01-OCT-2002 (TREMBlrel. 22, last annotation update)
DE	Surface antigen (Fragment).
OS	Trypanosoma cruzi.
OC	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX	NCB1_Taxid=5693;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CAI;
RX	MEDLINE=92389985; PubMed=1518528;
RA	Buschlazzo A., Campeella O.E.;
RT	"Sequence of the gene for a Trypanosoma cruzi protein antigenic during
RT	the chronic phase of human Chagas disease.";
RL	Mol. Biochem. Parasitol. 54:125-128(1992).
DR	EMBL; M92047; AAB97874.1; -.
FT	NON TER
SQ	SEQUENCE 722 AA; 74874 MW; C572189CD6DAB3E CRC64;
Query Match	16.5%; Score 194.5; DB 5; Length 722;
Best Local Similarity	29.8%; Pred. No. 5.9e-06;
Matches	59; Conservative 26; Mismatches 74; Indels 39; Gaps 6
Dy	20 PPFIEDNDODGHPHSPSINIPYGRNLPPPYRVPTVP-----SYGNVTYTD 70
Db	189 FYFGAYDEOLSRKHATIANVF-----LYNPLDAETGALANANKVSLP---FTBK 237
Dy	71 GLPSYPWILTSDFPYVVYHIRGPIPLATOLNVLPDPGRFPVPPSRFFSAAAAAPAPPIA 130
Db	238 -KPAASTATSPSVSHV-----TTVPATB-----PKSAEPKSAVPKSAEPKS 278
Dy	131 AEPAAPAPLTATPVAAEPPAGAPVAAEPALEAPVGAEPAAEPVAAEPVGEVPA 190
Db	279 AVPKSAEPKSAEPKSAEPKSAVPKSAEPKSAVPKSAEPKSAEPKSAEPKSAEPKS 338
Dy	191 EEPSPAEPATAKPAPEP 208
Db	339 AEPKSAEPKSAVPKSAEP 356
RESULT 11	
ID	C39266 PRELIMINARY; PRT; 3534 AA.
AC	C39266;
DT	01-JAN-1998 (TREMBlrel. 05, Created)
DT	01-JAN-1998 (TREMBlrel. 05, last sequence update)

DT	01-MAR-2003 (Tremblrel. 23, last annotation update)
DE	Counterpart of HSV-1 gene UL36 and VZV gene 22.
GN	24.
OS	Equine herpesvirus 4.
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC	Alphaherpesvirinae; Varicellovirus.
OX	NCBI_TaxID=10331;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NS80567;
RX	MEDLINE=96264497; PubMed=9603335;
RA	Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davidson A.J.;
RT	"The DNA sequence of equine herpesvirus-4.";
RL	J. Gen. Virol. 79:1197-1203 (1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NS80567;
RA	Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davidson A.J.;
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF030027; AAC59539.1; -
DR	InterPro; IPR006928; Herpes teg_N.
DR	InterPro; IPR005210; Herpes UL36.
DR	InterPro; IPR002965; P rich_extensn.
DR	Pfam; PF04843; Herpes teg_N_1.
DR	Pfam; PF03586; Herpes_UL36_1.
DR	PRINTS; P01217; PRICHEXTENSN.
SQ	SEQUENCE 3534 AA; 379956 MW; AF5164843CAlBFC6 CRC64;
Query Match	16.4%; Score 193; DB 12; Length 3534;
Best local similarity	37.4%; Pred. No. 3.5e-05;
Matches	70; Conservative 22; Mismatches 67; Indels 28; Gaps 12;
CY	36 PSLNIPYGIRNLPPPLYRPNVTVPYSYGGNTYTDTGLPSYPWILTSFGFPVYHIGRPL 95
DB	2697 PGYSIF--MOGLPDPDDNEALTLTAAPSKP---AAAPAPSKPAAPAPSKP-----AABA 2745
CY	96 ATQLNVPLLPKRGPFPPSPSFSSAAAAAPAAPFLAABP----AAALPTATPVNA-----E 147
DB	2746 PSKRAAAPAPSKPAAPAPAPSKP---PAAAPAPSKPAAPAPSKPAAPAPSKPAAPAPSK 2802
CY	148 PAAGAPVAABEPAAAPGAEPAAPVAABEPAAAPGVGEPAABEPSAPBPATA-KPAAP 206
DB	2803 PAA-APAPSKPA-AAPAPSKPA-AAPAPSKPA-AAPAPSKPA-AAPAPSKPAAPAPSKP 2857
CY	207 EPHPSPS 213
DB	2858 AAAAPAPS 2864
RESULT 12	
O49542	PRELIMINARY; PRT; 1108 AA.
ID	O49542.
AC	O49542.
DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE	Ciliary adhesion protein p97 (Cilium adhesin p97).
GN	p97.
OS	Mycoplasma hyopneumoniae.
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX	NCBI_TaxID=2099;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=232;
RX	MEDLINE=97175562; PubMed=9023217;
RA	Hsu T., Atkinson S., Minion F.C.;
RT	"Cloning and functional analysis of the p97 swine cilium adhesion gene
RT	of Mycoplasma hyopneumoniae.";
RL	J. Bacteriol. 179:1317-1323 (1997).
RN	[2]
RP	SEQUENCE OF 1089-1108 FROM N.A.
RC	STRAIN=232A;
RA	Hsu T., Minion F.C.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U50901; AAB47806.1; -
 DR EMBL; AF012905; AAC32526.1; -
 DR InterPro; IPR000719; Prot_kinase.
 DR PROSITE; PS00107; PROTEIN_KINASE_ABP; 1.
 SQ SEQUENCE 1108 AA; 124903 MW; 239CF62D61E11FE7 CRC64;

Query Match 16.0%; Score 188.5; DB 2; Length 1108;
 Best Local Similarity 51.2%; Pred. No. 2,4e-05;
 Matches 42; Conservative 8; Mismatches 27; Indels 5; Gaps 1;

QY 124 PAAPPIAEPAAAPLPTATPVAAPAGAPVAAEPAAEPVGAEPAAEPAAEPAAEP 183
 DB 813 PAAPPAAPVAAKPEPTTTPVAAKPE-----AAKPAAPVAAKPAAPVAAKPAAP 667
 QY 184 VGEPAAEPPSPAEPAAPAA 205
 DB 868 VAAKPEAPVAAKPAAPVAA 889

RESULT 13

ID Q9RX57 PRELIMINARY; PRT; 839 AA.
 AC Q9RX57;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein DR0458.
 GN DR0458.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 NC NCB1_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 Vamathevan J.J., Lam P., McDonald L., Uteerback T., Zalewski C.,
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 Fraser C.M.;
 RA "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RT Science 286:1571-1577(1999).
 DR EMBL; AE001904; AAF10038.1; -
 DR TIGR; DR0458; -
 DR InterPro; IPR002965; P rich_extensn.
 DR PRINTS; PR01217; PRICHTXTNSN.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 839 AA; 79759 MW; 3B6C2CD0CCEFFADE CRC64;

Query Match 16.0%; Score 188; DB 16; Length 839;
 Best Local Similarity 33.0%; Pred. No. 2e-05;
 Matches 64; Conservative 14; Mismatches 92; Indels 24; Gaps 9;

QY 27 DNDGDGHLPSILNIPIGIRNLPPPLTYRPNVTPSYPGNTYTDGSPSYWILTSP-GPP 85
 DB 140 DPTPAPPLKRP-----VQDTPPPVTPKPTPEPVTPKAPTPPEVLAQPPVAGTPPVAKP 193
 QY 86 YYTHIRGFPLAT-QLVNVPPLPPRGFPVPPSRFFSAAAAAPPIAAEPAAAPLTATPV 144
 DB 194 PV-----PAFTSQTPPTPVQAPARTPTPQPQ--AAAPFNAPAQTPAPATQAPAAQTPT 244
 QY 145 AAEBAAGAPVAAEPAAEPVGAEPAAEPVAAEPAAEPVGAEPAAEPVGAEPAAEPVGAEP 203
 DB 245 AQAAPATQTPPTPAAPAAQRPAGAP---SPAAPAAQANAPAGSVVPEATVPESTPA--P 299
 QY 204 AAPEHPSPSLBOA 217
 DB 300 SAQTP-PTPTRETA 312

RESULT 14

ID Q9RV74 PRELIMINARY; PRT; 395 AA.
 AC Q9RV74;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein DR0938.
 GN DR0938.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 NC NCB1_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 Vamathevan J.J., Lam P., McDonald L., Uteerback T., Zalewski C.,
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 Fraser C.M.;
 RA "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RT Science 286:1571-1577(1999).
 DR EMBL; AE001946; AAF10516.1; -
 DR TIGR; DR0938; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 395 AA; 40003 MW; 75FD96371DDAEP0 CRC64;

Query Match 15.8%; Score 186.5; DB 16; Length 395;
 Best Local Similarity 35.0%; Pred. No. 1.2e-05;
 Matches 57; Conservative 11; Mismatches 64; Indels 31; Gaps 9;

QY 60 PSYV-GNTYTDGSLSPYILITSPGPYVYHNRGFLATQLVNVPPLPPRGFPVPPSRFF 118
 DB 53 PAEPVGAPOPTTSTIPSPAVVADSPQT---AAPAQQTAGKIPV--DAAP----- 102
 QY 119 SAAPAAAP-----IAEPAAAPLTATPVAAPAGAPVAAEPAAEPVGAEPAAEP 170
 DB 103 ----APKTPPVLPBPRKMPAPPTPRVVEETTTAAPTQGPVQAAPATQPTPTPTPA 158
 QY 171 EAPVAAEPAAEPVGEVPA--EPPSPA---EPAT-AKPAAPP 208
 DB 159 QAPATQPPATPAP---EPAPAAEPAPATPPEBAPTPEPAAP 198

RESULT 15

ID Q99K31 PRELIMINARY; PRT; 616 AA.
 AC Q99K31;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to procollagen, type VI, alpha 3 (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005491; AAH05491.1; -
 DR HSSP; P12111; IKT.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR002965; P rich_extensn.
 DR InterPro; IPR002035; VWF_A.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:06:54 : Search time 18 seconds

(without alignments)
572.158 Million cell updates/sec

Title: US-09-923-236-2

Sequence: 1 MKLLLMACIVCVAFARAKRRF.....TAKPAPRPHPSPLEQANQ 219

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	180	15.3	555	1	GPI_CHLRE
2	175	14.9	1054	1	IF2_STRAU
3	175	14.9	3149	1	TBGT_BBV
4	174	14.8	518	1	TPW4_DROME
5	170	14.4	3176	1	CA36_HUMAN
6	169.5	14.3	802	1	ENAH_MOUSE
7	169	14.3	299	1	RL22_DROME
8	169	14.3	865	1	CPN_DROME
9	167.5	14.2	450	1	CYL_PARDE
10	166	14.1	810	1	NFM_BOVIN
11	165.5	14.0	439	1	XP2_XENLA
12	162	13.8	5147	1	PCLO_HUMAN
13	160.5	13.6	992	1	EBN6_EBV
14	158	13.4	1274	1	ENAH_MOUSE
15	154	13.1	353	1	CCPA_ACEXY
16	154	13.1	4499	1	DYHA_CHIRE
17	152.5	12.9	5038	1	PCLO_MOUSE
18	151	12.8	721	1	YK82_MYCTU
19	151	12.8	88	1	H82_NEIGO
20	151	12.8	211	1	TUBB_SOLATU
21	150	12.7	352	1	ALGP_PSEAE
22	148.5	12.6	464	1	S3A2_HUMAN
23	147	12.5	1386	1	ZAP2_MOUSE
24	146.5	12.4	477	1	MAZ_MOUSE
25	145.5	12.4	361	1	IF35_MOUSE
26	145.5	12.4	1565	1	PAC_STRMU
27	145	12.3	475	1	S3A2_MOUSE
28	145	12.3	2167	1	SHK1_RAT
29	144.5	12.3	477	1	MAZ_HUMAN
30	142.5	12.1	497	1	WAS2_HUMAN
31	142.5	12.1	1083	1	T2D3_HUMAN
32	142	12.1	2161	1	SHK1_HUMAN
33	141.5	12.0	316	1	CDNC_HUMAN

34	141.5	12.0	465	1	PXD1_HUMAN
35	140.5	11.9	852	1	WS14_HUMAN
36	140	11.9	3421	1	TBGT_HSVB
37	140	11.9	5085	1	PCLO_RAT
38	139.5	11.8	397	1	SEPL_MOUSE
39	139.5	11.8	1157	1	BBC1_YEAST
40	138.5	11.8	206	1	PRP1_MEDTR
41	138.5	11.8	1183	1	DRPL_RAT
42	138	11.7	376	1	FXL2_HUMAN
43	138	11.7	2517	1	NCR2_HUMAN
44	137	11.6	236	1	PRP_MEDSA
45	136.5	11.6	1003	1	MBD6_HUMAN

ALIGNMENTS

RESULT 1
GPI_CHLRE
ID GPI_CHLRE STANDARD; PRT; 555 AA.
AC Q9FQ6; Q03927; (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2115092; PubMed=11258910;
RA Ferris P.J., Woessner J.P., Walfenrichmidt S., Kitz S., Drees J., Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in plant hydroxyproline-rich glycoproteins";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC - FUNCTION: Major component of the outer cell wall w6 (crystalline) layer.
CC - SUBUNIT: Associates with GP2 and GP3.
CC - PTM: N-glycosylated and O-glycosylated.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC EMBL, AF309494; AAC45420.1; -
CC EMBL, M58496; AAB69706.1; ALT_SEQ.
CC Glycosylated; Q9FQ6; -
CC InterPro: IPR002965; P-rich extensin.
CC InterPro: IPR003882; Plectin_extensin.
CC PRINTS: PRO1217; PRICHEXTENSIN.
CC PRINTS: PRO1218; PSTLTEXTENSIN.
CC Glycoprotein; Repeat; Signal.
CC SIGNAL 1 29
FT CHAIN 30 555
FT DOMAIN 40 339
FT DOMAIN 259 339
FT CARBOHYD 339 399
FT CARBOHYD 455 455
FT POTENTIAL.
FT VEGETATIVE CELL WALL PROTEIN GPI.
FT 49 X 5 AA APPROXIMATE PSEPX REPEATS.
FT POLY-PRO.
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455


```

Db      290  PPVWLPEFRRIITPYVLRPLPSKSTFSDSEFPARKYSPAKTNSDP--SSPASADPASAAPAS 347
Qy      146  AEPAGAEEVAAEPAAEAEPVGAEPAAEAEPVAAEPAAEAEPVGAEPAAEAEPSPAPAPATAKP-A 204
Db      348  AAPASAAAPASAAAPASAAAPASAAAPASAAAPASPPLPITPIGLGHTGTGVAFTSPAPASGA 407
Qy      205  APE-----PHSPS 213
Db      408  APQTPKRRKGLGKDSFHKKFT 428

RESULT 4
TPM4_DROME STANDARD; PRT; 518 AA.
AC      P49455; P49456; Q24425; Q24426;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      TROPOMYOSIN I, Isoforms 33/34 (Tropomyosin II).
GN      TM1 OR TM11.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC      TISSUE=Embryo, and Pupae;
RX      MEDLINE=89127197; Pubmed=2851721;
RA      Hanke P.D., Scotti R.V.;
RT      "The Drosophila melanogaster tropomyosin II gene produces multiple
RT      proteins by use of alternative tissue-specific promoters and
RT      alternative splicing."
RT      Mol. Cell. Biol. 8:3591-3602(1988).
RL      [2]
RP      SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC      STRAIN=Oregon-R; TISSUE=Pupae;
RX      MEDLINE=87064486; Pubmed=3097506;
RA      Karlik C.C., Fyberg E.A.;
RT      "Two Drosophila melanogaster tropomyosin genes: structural and
RT      functional aspects."
RL      Mol. Cell. Biol. 6:1965-1973(1986).
RN      -1- ALTERNATIVE PRODUCTS:
RP      Event=Alternative splicing; Named isoforms=5;
RC      Comment=Additional isoforms seem to exist;
RX      Name=33; Synonyms=9c;
CC      IsoId=P49455-1; Sequence=Displayed;
CC      Name=Muscle; Synonyms=9D;
CC      IsoId=P06754-1; Sequence=External;
CC      Name=Non-muscle; Synonyms=Cytoskeletal;
CC      IsoId=P06754-2; Sequence=External;
CC      Name=9A;
CC      IsoId=P06754-3; Sequence=External;
CC      Name=34; Synonyms=9B;
CC      IsoId=P49455-2; Sequence=VSP 006623, VSP 006624, VSP 006625;
CC      -1- TISSUE SPECIFICITY: Both isoforms are only expressed in indirect
CC      flight muscles
CC      -1- DEVELOPMENTAL STAGE: Both isoforms are expressed during pupal and
CC      adult stages.
CC      -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC      EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC      -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X76208; CAA53800.1; -.

```

Query Match	Best Local Similarity	Matches	Score 174;	DB 1;	Length 518;
17	119	114	30.28;	Pred. No. 0.0018;	
264	183	183	13;	Mismatches 107;	Indels 58; Gaps 7
59	199	199			
324	503	503			
92	518	54558			
384	518	54558			
149	518	54558			
444	518	54558			
199	518	54558			

DB 504 AAAAAAAAAA 518

RESULT 5
CA36 HUMAN STANDARD; PRT; 3176 AA.
ID P12111; Q16501;
AC 01-OCT-1989 (Rel. 12, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3 (VI) chain precursor.
GN COL6A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=90151612; PubMed=1689238;
RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,
RA Glavville R., Mayer U., Mann K., Deutzmann R., Timpl R.;
RT "Mosaic structure of globular domains in the human type VI collagen
RT alpha 3 chain: similarity to von Willebrand factor, fibronectin,"
RT actin, salivary proteins and apocytin type protease inhibitors,"
RT EMBO J. 9:385-393(1990).
RN [2]
RP REVISIONS.
RA Chu M.-L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 2038-2373 FROM N.A.
RX MEDLINE=89066644; PubMed=3198591;
RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,
RA Timpl R.;
RT "Amino acid sequence of the triple-helical domain of human collagen
RT type VI,"
RT J. Biol. Chem. 263:18601-18606(1988).
RN [4]
RP SEQUENCE OF 2092-2157 FROM N.A.
RX MEDLINE=88029444; PubMed=3665927;
RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,
RA Han-Chen C.-C., Bernard M.P., Timpl R.;
RT "Characterization of three constituent chains of collagen type VI by
RT peptide sequences and cDNA clones,"
RT Eur. J. Biochem. 168:309-317(1987).
RN [5]
RP SEQUENCE OF 2092-2151 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88161046; PubMed=3348212;
RA Weill D., Mettel M.-G., Passage E., van Cong N., Pribula-Conway D.,
RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal localization of human gene encoding the
RT three chains of type VI collagen,"
RT Am. J. Hum. Genet. 42:435-445(1988).
RN [6]
RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=93054780; PubMed=1339440;
RA Zanussi S., Doliana R., Segat D., Ronaldo P., Colombatti A.;
RT "The human type VI collagen gene. mRNA and protein variants of the
RT alpha 3 chain generated by alternative splicing of an additional 5-end
RT exon,"
RT J. Biol. Chem. 267:24082-24089(1992).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.
RX MEDLINE=95182468; PubMed=7533217;
RA Arnoux B., Merigues K., Saludjian P., Norris F., Norris K., Bjoern S.,
RA Olsen O., Petersen L., Ducruix A.;
RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of
RT human type VI collagen,"
RT J. Mol. Biol. 246:609-617(1995).
RN [8]

RP STRUCTURE BY NMR OF 3102-3164.
RX MEDLINE=96398604; PubMed=8805527;
RA Zweckstetter M., Cziech M., Mayer U., Chu M.-L., Zinth W., Timpl R.,
RA Holak T.A.;
RT "Structure and multiple conformations of the Kunitz-type domain from
RT human type VI collagen alpha3 (VI) chain in solution,"
RT Structure 4:195-209(1996).
RN [9]
RP STRUCTURE BY NMR OF 3107-3164.
RX MEDLINE=97410331; PubMed=9265624;
RA Soerensen M.D., Bjoern S., Norris K., Olsen O., Petersen L.,
RA James T.L., Led J.J.;
RT "Solution structure and backbone dynamics of the human alpha3-chain
RT type VI collagen C-terminal Kunitz domain,"
RT Biochemistry 36:10439-10450(1997).
RN [10]
RP DISEASE.
RX MEDLINE=21987636; PubMed=11992252;
RA Demir E., Sabatelli P., Allamand V., Ferreira A., Moghadaszadeh B.,
RA Makrelouf M., Topaloglu H., Echenne B., Merlini L., Guicheney P.;
RT "Mutations in COL6A3 cause severe and mild phenotypes of Ullrich
RT congenital muscular dystrophy,"
RT Am. J. Hum. Genet. 70:1446-1458(2002).
RN [11]
RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.
RX MEDLINE=98204804; PubMed=9536084;
RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,
RA Stajich J.W., Viles K., Vance J.W., Chu M.-L., Speer M.C.;
RT "Missense mutation in a von Willebrand factor type A domain of the
RT alpha 3 (VI) collagen gene (COL6A3) in a family with Bethlem
RT myopathy,"
RT Hum. Mol. Genet. 7:807-812(1998).
RN [12]
RP FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (VI),
CC ALPHA 2 (VI), AND ALPHA 3 (VI).
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1;
CC IsoId=P12111-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P12111-2; Sequence=VSP 001172;
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- DISEASE: Defects in COL6A3 are a cause of Bethlem myopathy (BM)
CC [MIM:158810]. BM is a rare autosomal dominant proximal myopathy
CC characterized by early childhood onset (complete penetrance by the
CC age of 5) and joint contractures most frequently affecting the
CC elbows and ankles.
CC -1- DISEASE: Defects in COL6A3 are a cause of Ullrich congenital
CC muscular dystrophy (UCMD) [MIM:254090]; also known as Ullrich
CC scleroatonic muscular dystrophy, an autosomal recessive congenital
CC myopathy. UCMD is characterized by muscle weakness and multiple
CC joint contractures, generally noted at birth or early infancy. The
CC clinical course is more severe than in Bethlem myopathy.
CC -1- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- SIMILARITY: Contains 12 WFPA domains.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcel/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X52022; CAA36267.1; -;
CC EMBL; X06186; CAA29957.1; -;
CC EMBL; M20778; -; NOT_ANNOTATED_CDS.
CC EMBL; M27449; AAB52057.1; -;
CC EMBL; S49432; AAB24261.1; -;
CC PIR; AS9140; CGHJ3A.
CC PDB; 1KNT; 01-NOV-94.

DR EMBL, U72520; AAC52863.1; -
 DR EMBL, U72521; AAC52864.1; -
 DR EMBL, U72522; AAC52865.1; -
 DR EMBL, U72523; AAC52866.1; -
 DR PIR, S27200; S27200.
 DR MGI, MGI:108360; Bnah.
 DR InterPro, IPR001960; WH1.
 DR Pfam, PF00568; WH1; 1.
 DR SMART, SM00461; WH1; 1.
 DR Developmental protein; Neurogenesis; Alternative splicing.
 FT DOMAIN 1 108
 FT DOMAIN 442 464
 FT DOMAIN 542 552
 FT DOMAIN 562 574
 FT DOMAIN 578 589
 FT DOMAIN 593 605
 FT VARSPLIC 1 412
 FT VARSPLIC 535 631
 FT VARSPLIC 117 131
 FT VARSPLIC 132 135
 FT VARSPLIC 117 135
 FT VARSPLIC 259 500
 FT VARSPLIC 802 AA; 85844 MW; 5928B975E20F77F CRC64;
 SQ SEQUENCE 802 AA; 85844 MW; 5928B975E20F77F CRC64;
 Query Match 14.4%; Score 169.5; DB 1; Length 802;
 Best Local Similarity 27.8%; Pred. No. 0.0064;
 Matches 64; Conservative 22; Mismatches 91; Indels 53; Gaps 10;
 QY 31 GHPHPSLNIPIYGIKRLPPPLYYR-----PVNTVPS-----YPGNTYDTGLPSYP 76
 DB 330 GSAFHVLT--PH-YATVPRPLNKNRSPSVNTSSQPPAKSCAWPTSNFSLP-PSP 385
 QY 77 WLITSGFVYVHIRPPLATQLNVPPLPR-----GFPVPSRFRSAAAAA 125
 DB 386 MISSPGRATGRPVLPVCSVPQMPSPPTAPNGSLDSVTYPPSPPT--SGPAAP 443
 QY 126 APPIAEPPAAAPLTATPV-----AAEPAGAPVAAEPAAEAPVGAEPAAEAPV 174
 DB 444 PPPPPPPPPPLPPLPPLPLASLSHCQSQASPPFGTPLASTPSKPSVLPSPSAGATA 503
 QY 175 AAEPAAEAPV-----VEPA-----AEEPSPAEPATAPAAEPPSPS 213
 DB 504 SAETPLNPPELGDSASEPGLQAASQAESEPTFGVLGPPAPPPPLPS 553
 RESULT 7
 ID RL22 DROME STANDARD; PRT; 299 AA.
 AC P50887; O9VX9;
 DT 01-NOV-1997 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 608 ribosomal protein L22.
 GN RPL22 OR EG:BACR19J1.4 OR CG7434.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI TaxID=7227;
 OX [1]
 RP SEQUENCE FROM N.A.
 RA Glover C.V.C., Bidwai A.P., Zhao W.F.;

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196006; PubMed=10731137;
 RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demalbert J., Borikova D.,
 RA Dreano S., Gloux S., Laure V., Mottier S., Galibert F., Borikova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kimios I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madeno E., Cadieu E.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourikot F.,
 RA Beinert N., Dowe G., Schaefer U., Jaackle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlancou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.,
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222 (2000).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George K.Y., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Chape M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayy A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Botchko S.,
 RA Borikova D., Borikova M.R., Bouck J., Brockhead P., Brotler P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Dey S., Davis S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mewlrow G., Milne N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 CC - SIMILARITY: BELONGS TO THE L22E FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL, U42587; AAB17433.1; -
 CC DR EMBL, AL132792; CAB60023.1; -
 DR EMBL, AB003418; AAF45546.1; -

```

DR  FLYBase: FBgn0015288; RPL22.
DR  InterPro: IPR002671; Ribosomal_L22e.
DR  Pfam: PF01776; Ribosomal_L22e; 1.
DR  ProDom: PD007306; Ribosomal_L22e; 1.
KW  Ribosomal protein.
FT  DOMAIN 24 31 POLY-ALA.
FT  DOMAIN 46 50 POLY-ALA.
FT  DOMAIN 65 70 POLY-ALA.
FT  DOMAIN 93 98 POLY-ALA.
FT  DOMAIN 103 112 POLY-ALA.
FT  DOMAIN 136 152 POLY-ALA.
FT  DOMAIN 185 188 POLY-LYS.
FT  DOMAIN 292 299 ASP/GLU-RICH (HIGHLY ACIDIC).
SO  SEQUENCE 239 AA; 30610 MW; 46A99005610EAEB0 CRC64;

Query Match 14.3%; Score 169; DB 1; Length 239;
Best Local Similarity 44.1%; Pred. No. 0.0022;
Matches 49; Conservative 6; Mismatches 34; Indels 22; Gaps 5

Cc  FLYBase: FBgn0015288; RPL22.
Cc  InterPro: IPR002671; Ribosomal_L22e.
Cc  Pfam: PF01776; Ribosomal_L22e; 1.
Cc  ProDom: PD007306; Ribosomal_L22e; 1.
Cc  Ribosomal protein.
Cc  DOMAIN 24 31 POLY-ALA.
Cc  DOMAIN 46 50 POLY-ALA.
Cc  DOMAIN 65 70 POLY-ALA.
Cc  DOMAIN 93 98 POLY-ALA.
Cc  DOMAIN 103 112 POLY-ALA.
Cc  DOMAIN 136 152 POLY-ALA.
Cc  DOMAIN 185 188 POLY-LYS.
Cc  DOMAIN 292 299 ASP/GLU-RICH (HIGHLY ACIDIC).
Cc  SEQUENCE 239 AA; 30610 MW; 46A99005610EAEB0 CRC64;

RESULT 8
CPN_DROME
ID  CPN_DROME STANDARD; PRT; 865 AA.
AC  Q02910;
DT  01-OCT-1993 (Rel. 27, Created)
DT  01-OCT-1993 (Rel. 27, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  Calphotcin.
GN  CPN OR CAP.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC  Empidoidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
[1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=Canton-S;
RX  MEDLINE=93165729; PubMed=8094559;
RA  Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT  "Calphotcin: a Drosophila photoreceptor cell calcium-binding protein.";
RL  Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
[2]
RN  SEQUENCE FROM N.A.
RP  STRAIN=Canton-S;
RX  MEDLINE=93165730; PubMed=8434015;
RA  Ballinger D.G., Xue N., Harshman K.D.;
RT  "A Drosophila photoreceptor cell-specific protein, calphotcin, binds
RT  calcium and contains a leucine zipper.";
RL  Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
-1- FUNCTION: Might function as a calcium-sequestering "sponge" to
Cc  regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
Cc  of Ca(2+) per mole of protein.
Cc  -1- SUBUNIT: Homodimer (Probable).
Cc  -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
Cc  -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
Cc  COMPOUND EYES AND OCELLI.
Cc  -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
Cc  DEVELOPMENT.
Cc  -----
Cc  This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc  between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc  the European Bioinformatics Institute. There are no restrictions on its
Cc  use by non-profit institutions as long as its content is in no way
Cc  modified and this statement is not removed. Usage by and for commercial
Cc  entities requires a license agreement (See http://www.isb-sb.ch/announce/

```

```

CC or send an email to license@lab-stb.ch).
CC -----
CC DR EMBL, L02111; AAA28405.1; -.
CC DR EMBL, L05080; AAA28420.1; -.
CC DR PIR; A47282; A47282.
CC DR PIR; A47283; A47283.
CC DR FlyBase; FBpp0010218; Cpn.
CC DR GO; 00005509; F:calcium ion binding activity; IDA.
CC KW Calcium-binding.
CC FT CONFLICT 36 A -> AAVAPAVA (IN REF. 2) .
CC FT CONFLICT 43 43 I -> T (IN REF. 2) .
CC FT CONFLICT 64 64 I -> V (IN REF. 2) .
CC FT CONFLICT 76 76 T -> A (IN REF. 2) .
CC FT CONFLICT 100 100 P -> P (IN REF. 2) .
CC FT CONFLICT 126 127 VQ -> AP (IN REF. 2) .
CC FT CONFLICT 154 154 I -> V (IN REF. 2) .
CC FT CONFLICT 160 160 S -> T (IN REF. 2) .
CC FT CONFLICT 534 534 A -> E (IN REF. 2) .
CC FT CONFLICT 699 699 I -> T (IN REF. 2) .
CC FT CONFLICT 703 703 V -> L (IN REF. 2) .
CC FT CONFLICT 721 721 D -> E (IN REF. 2) .
CC SQ SEQUENCE 865 AA; 84781 MM; 2110417BOB087CFE CRC64;
CC
CC Query March 14.3%; Score 169; DB 1; Length 865;
CC Best Local Similarity 29.5%; Pred. No. 0.0053;
CC Matches 70; Conservative 14; Mismatches 75; Indels 78; Gaps 11
CC
CC QY 55 PNTVPSYGTNTYDTGLPSYPMILTSBGFVYVHINRFPPLAT--QLNVPELPGRGPF 111
CC Db 42 PIATVPAPPTTASV-----QPATVTPAPAPIAASTVPASVASVPVAAPTPPASVPV 97
CC QY 112 VPP--SRFSAAPAAPAPPIAEPDA-----AAAPLTAT-PPAA-----PAA----- 150
CC Db 98 STPVAVAQIPVAVSAPVAPVAPVAPVATPTPVQIPVAPVATPTPVVAAASAPVATPVVISV 157
CC QY 151 -----GAPVAAEP-----AAPVGAE-----PAAAPVA 175
CC Db 158 IASPPVVPANTTVPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 217
CC QY 176 AEPAAEPV-----GEPAAEBSPPAPATKAPPAEPHPS--PSLEQA 217
CC Db 218 TTFPCVAPLIEVSVAATKPLAAAEFPVVAAPATETPPVAPAAASPHVSVAPAVETA 274
CC
CC RESULT 9
CC CY1_PARDE STANDARD; PRT; 450 AA.
CC ID CY1_PARDE
CC AC P15627;
CC DT 01-JAN-1990 (Rel. 13, Created)
CC DT 01-JAN-1990 (Rel. 13, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Cytochrome c1 precursor.
CC GN PUTC.
CC OS Paracoccus denitrificans.
CC OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
CC OC Rhodobacteraceae; Paracoccus.
CC OX NCB1_TaxID=266;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=88007612; PubMed=2820981;
CC RA Kuroweki B., Ludwig B.;
CC RT "The genes of the Paracoccus denitrificans bcl complex. Nucleotide
CC sequence and homologs between bacterial and mitochondrial
CC subunits ";
CC RL J. Biol. Chem. 262:13805-13811 (1987).
CC CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS. C1 FUNCTIONS AS AN ELECTRON DONOR TO
CC CYTOCHROME C.
CC CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIBESKE PROTEIN.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M17522; AAA2573.1; -
DR EMBL; X05799; CAA29245.1; -
DR PIR; C29413; C29413.
DR InterPro; IPR0002326; Cyt C1.
DR InterPro; IPR000345; CytC_heme_bind.
DR Pfam; PF02167; Cytochrome_C1.
DR PRINTS; PR00603; CYTOCHROME_C1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Electron transport; Respiratory chain; Heme; Transmembrane; Signal.
FT SIGNAL 1 21
FT CHAIN 1 21
FT BINDING 22 450 CYTOCHROME C1.
FT BINDING 245 245 HEME (COVALENT).
FT BINDING 248 248 HEME (COVALENT).
FT METAL 249 249 IRON (HEME AXIAL LIGAND).
FT METAL 373 373 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT TRANSMEM 421 435 ANCHORS TO THE MEMBRANE (POTENTIAL).
SQ SEQUENCE 450 AA; 46874 MW; C1EA8BE48FE7603 CRC64;

Query Match 14.2%; Score 167.5; DB 1; Length 450;
Best Local Similarity 42.6%; Pred. No. 0.0037;
Matches 58; Conservative 8; Mismatches 35; Indels 35; Gaps 8;

QY 114 PRRFSAAPAAAPPIAAEPAA--APLATPVAA-----EPAGAPVA-----AEPAA 159
DB 48 PAAADTAPAAEAADPEAAEAAGKAEVTEBPATTPAEBPAAEPAAATEBPDAEABA 107
QY 160 -----AAEAVGAEPAAEAAPVAAEPAAEPAAEPAAEPAAEP--ATAKPA 205
DB 108 ABEAQATTEAPAEPEAAEPAAEP--AEBPADAA--AEBAAAEAPAEPEAAAEPPAA 164
QY 206 PEP---HPSPSLEQA 217
DB 165 EEPATEEAPAEPEAA 180

RESULT 10
NEW_BOVIN STANDARD; PRT; 810 AA.
AC 077788;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
OS NEF3 OR NEFM OR NFM.
GN Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hill W.D., Balin B.J., Sprinkle T.J., Spicer K.,
RA Gearhart D.A.;
RT "The bovine neurofilament M subunit has a novel set of KSP repeats
RT normally restricted to NF-H.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER (BY SIMILARITY).
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING

CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF091342; AAC36357.1; -
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation.
FT NON_TER 1 1
FT DOMAIN <1 296 ROD.
FT DOMAIN 297 810 TAIL.
FT DOMAIN <1 20 COIL 1A.
FT DOMAIN 21 33 LINKER 1.
FT DOMAIN 34 132 COIL 1B.
FT DOMAIN 133 149 LINKER 12.
FT DOMAIN 150 171 COIL 2A.
FT DOMAIN 172 175 LINKER 2.
FT DOMAIN 176 296 COIL 2B.
FT DOMAIN 503 582 8 X 13 AA TANDEM REPEATS.
SQ SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;

Query Match 14.1%; Score 166; DB 1; Length 810;
Best Local Similarity 35.1%; Pred. No. 0.0073;
Matches 34; Conservative 12; Mismatches 51; Indels 0; Gaps 0;

QY 112 VPPSRFFSMAAPPAAPPIAAEPAAAPLITATPVAAEPAAAGAPVAAEPAAEPVGAEPAAE 171
DB 484 VAPKEELAAEAQVKEPEKAKSPVAPSPPTKSPAKSPKAPKSPKSPKSPKSPKSPK 543
QY 172 APVAAEPAAEPVGAEPAAEPSPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 208
DB 544 SPKAPSPKAPSPKAPSPKAPSPKAPSPKAPSPKAPSPKAPSPKAPSPKAPSPKAPSP 580

RESULT 11
XP2_XENLA STANDARD; PRT; 439 AA.
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-1995 (Rel. 42, Last annotation update)
DE Skin secretory protein xp2 precursor (APBG protein).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).
RX MEDLINE=92332564; PubMed=1629230;
RA Hauser F., Roeben C., Hoffmann W.;
RT "XP2, a new member of the P-domain peptide family of potential growth
RT factors, is synthesized in Xenopus laevis skin.";
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).
RX TISSUE=Skin;
RC MEDLINE=90127399; PubMed=2298293;
RA Gmachl M., Berger H., Thalhammer J., Krell G.;
RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly

RT repetitive amino acid sequence.";
 RL FEBS Lett. 260:145-148(1990).
 CC -1- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE
 CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS
 CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=APRG;
 CC IsoId=PI7437-1; Sequence=Displayed;
 CC Name=2; Synonyms=XP2;
 CC IsoId=PI7437-2; Sequence=VSP_004652;
 CC -1- TISSUE SPECIFICITY: SKIN.
 CC -1- SIMILARITY: Contains 2 P-TYPE (trefoil) domains
 CC -1- CAUTION: REP 2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392
 CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, M90095; AAA50001.1; -
 CC DR EMBL, X51394; CAA35759.1; ALT_FRAME.
 CC DR PIR, A37331; A37331.
 CC DR HSSP, P04155; 1P82.
 CC DR InterPro, IPR000519; P_trefoil.
 CC DR Pfam, PF00088; trefoil.2.
 CC DR PRINTS, PR00680; PTFREOIL.
 CC DR SMART, SM00018; PD; 2. TREFOIL.
 CC DR PROSITE, PS00025; P_TREFOIL, 2.
 CC KW Signal; Growth factor; Alternative splicing; Repeat;
 CC KM Pyrolydione carboxylic acid.
 CC FT SIGNAL 1 22 POTENTIAL.
 CC FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
 CC FT MOD_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
 CC FT DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP] (2,4)-A-E.
 CC FT DOMAIN 350 391 P-TYPE 1.
 CC FT DOMAIN 397 438 P-TYPE 2.
 CC FT DISULFID 351 377 BY SIMILARITY.
 CC FT DISULFID 361 376 BY SIMILARITY.
 CC FT DISULFID 371 388 BY SIMILARITY.
 CC FT DISULFID 398 424 BY SIMILARITY.
 CC FT DISULFID 408 423 BY SIMILARITY.
 CC FT DISULFID 418 435 BY SIMILARITY.
 CC FT VARSPLC 26 343 Missing (in isoform 2).
 CC FT CONFLICT 3 3 FTId=VSP_004652.
 CC FT CONFLICT 18 18 H -> S (IN REF. 2).
 CC FT CONFLICT 18 18 C -> W (IN REF. 2).
 CC SQ SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;
 CC -----
 CC Query Match 14.0%; Score 165.5; DB 1; Length 439;
 CC Best Local Similarity 39.7%; Pred. No. 0.0047;
 CC Matches 48; Conservative 6; Mismatches 56; Indels 11; Gaps 3;
 CC -----
 CC QY 103 PLPPGPFVPPSPRFFSAAPAPAPPIAEPAPAAAPLTATPVAAEPAG-APVAEPAAE 161
 CC DB 103 PAPAGGAPAP-----AEGGAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAP 157
 CC QY 162 APVGAEPAAEPVAAEPAAEPVGVPAEPSPAE-----DATAKPAEPHPSPSLAQ 216
 CC DB 158 APAPAEVAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGE 217
 CC QY 217 A 217
 CC DB 218 A 218

ID PCLO HUMAN STANDARD; PRT; 5147 AA.
 AC Q9Y6V0; Q43373; Q60305; Q9BVCG; Q9U1V2; Q9Y6U9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Aczonin) (Fragments).
 GN PCLO OR ACZ OR KIA00559.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE OF 1-759 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99439764; PubMed=10508862;
 RA Wang X., Kibbechull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 RA Kiliann M.W.;
 RT "Aczonin, a 550-kD putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rlm and bassoon and binds
 RT profilin.";
 RL J. Cell Biol. 147:151-162(1999).
 RN [2]
 RP SEQUENCE OF 552-4404 FROM N.A.
 RA Kraemer J., Wollam C., Woldmann P., McGrane B.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 3619-5147 FROM N.A. (ISOPFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 RN [4]
 RP SEQUENCE OF 4405-4439 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Uscio T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Faney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalek U., Smalhe D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Werra M.A.;
 RT "Genetic and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Elliott G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
 CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;

RESULT 12
 PCLO_HUMAN

```

CC      IsoId=Q9Y6V0-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
CC      VSP_003926, VSP_003927;
CC      Note=No experimental confirmation available;
CC      -1- DOMAIN: C2 domain 1 is involved in binding calcium and
CC      phospholipids. Calcium binds with low affinity but with high
CC      specificity and induces a large conformational change.
CC      -1- SIMILARITY: Contains 2 C2 domains.
CC      -1- SIMILARITY: Contains 1 PDZ/DHR domain.
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y19188; CAB60727.1; -
CC      EMBL: AC049803; AAD20936.1; -
CC      EMBL: AC049886; AAD21789.1; -
CC      EMBL: AB011131; BAA25485.1; -
CC      EMBL: BC001304; AAB01304.1; -
CC      EMBL: AC004082; AAB97937.1; -
CC      HSP: P04410; T00634.
CC      HSP: P04410; IAZ5.
CC      GeneW: HGNC:13406; PCLO.
CC      MIM: 604918; -
CC      DR GO: GO:0005856; C:cytoskeleton; NAS.
CC      DR GO: GO:0045202; C:synaptic junction; ISS.
CC      DR GO: GO:0005509; F:calcium ion binding activity; ISS.
CC      DR GO: GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
CC      DR GO: GO:0005522; F:profilin binding activity; ISS.
CC      DR GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC      DR GO: GO:0016080; P:synaptic vesicle targeting; ISS.
CC      DR InterPro: IPR000008; C2.
CC      DR InterPro: IPR001565; Synapto-tagmin.
CC      DR PRINTS: PR00360; C2DOMAIN.
CC      DR PRINTS: PR00399; SYNAPTOTAGMIN.
CC      DR SMART: SM00239; C2; 2.
CC      DR PROSITE: PS00499; C2_DOMAIN_1; 1.
CC      DR PROSITE: PS00004; C2_DOMAIN_2; 2.
CC      KM Calcium/phospholipid-binding; zinc; Metal-binding; zinc-finger;
CC      Repeat; Alternative splicing.
CC      FT      NON_TER 1
CC      FT      DOMAIN 400 465
CC      FT      ZN_FING 499 523
CC      FT      ZN_FING 969 992
CC      FT      NON_CONS 1010 1011
CC      FT      DOMAIN 2300 2325
CC      FT      DOMAIN 4391 4442
CC      FT      DOMAIN 4544 4633
CC      FT      DOMAIN 5031 5121
CC      FT      VARSPLIC 4404 4404
CC      FT      VARSPLIC 4534 4534
CC      FT      VARSPLIC 4576 4576
CC      FT      VARSPLIC 4757 4761
CC      FT      VARSPLIC 4762 5147
CC      FT      SEQUENCE 5147 AA; 563537 MW; CD5D84990498CDD3C CRC64;
CC      Query Match 13.8%; Score 162; DB 1; Length 5147;
CC      Best Local Similarity 23.9%; Pred. No. 0.057;
CC      Matches 60; Conservative 31; Mismatches 78; Indels 82; Gaps 10;
CC      31 GPHLHSLNIPYGRNLPPPLYRPNVTVSPVSGNTYTDG---LPSYPIILTSPPGFV 87

```

```

Db      255 GESVKSPLSP-----SKPPIQPTGKPPAQPGHEKSGPPAKPPAPSGILTK----- 304
QY      88 YHIRFPLATQLNV--PLPLPRGF-----PVPSPRRFSA----- 120
Db      305 -----PLAQPGTVKPVQPPGTGTPKPPAPLGLGPKPPAQQTGSSEKSSSEDPGPPALAQ 358
QY      121 -----AAFAAP-----PIAEPAAAAPL-----TATPVAEPAAAGAPVAE 157
Db      359 PGVGTTPAQPGPPAPPPQGVGTPTKPLAQGGLGSPAPAPGPTTKPPAQTKPPSQGPGSTK 418
QY      158 PAAPAPVGAEPAAEPVAAEPAAEPVGAEPVGAEPAAEPAAEPAT-----AKPA 205
Db      419 PPPQPGAPKAPSPQPGSGTKPPSQGPGSAKPSAQGPSAPKPSAQGPTKPVSGTGFCKPLQ 478
QY      206 PEPHPSPLSQ 216
Db      479 P-PTVSPSAKQ 488

RESULT 13
ID      EBN6_EBV STANDARD; PRT; 992 AA.
AC      P03204;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-AUG-1988 (Rel. 08, Last sequence update)
DT      01-FEB-1994 (Rel. 28, Last annotation update)
DE      EBNA-6 nuclear protein (EBNA-3C) (EBNA-4B).
GN      BBRF3-BBRF4.
OS      Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Gammaherpesvirinae; Lymphocryptovirus.
OX      NCBI_TaxID=10377;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=84270667; PubMed=6087149;
RA      Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA      Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seglin C.,
RA      Tuffnell P.S., Barrell B.G.;
RT      "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RN      [2]
RP      Nature 310:207-211 (1984).
RN      [3]
RP      CHARACTERIZATION.
RX      MEDLINE=88155772; PubMed=2831394;
RA      Petri L., Sample J., Wang F., Kieff E.;
RT      "A fifth Epstein-Barr virus nuclear protein (EBNA3C) is expressed in
RT      latently infected growth-transformed lymphocytes.";
RN      [4]
RP      J. Virol. 62:1330-1338 (1988).
RN      [5]
RP      SUBCELLULAR LOCATION.
RX      MEDLINE=90266473; PubMed=2161150;
RA      Petri L., Sample C., Kieff E.;
RT      "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT      latent infection nuclear proteins.";
RN      [6]
RP      Virology 176:563-574 (1990).
RN      [7]
RP      -1- FUNCTION: INVOLVED IN LATENT CYCLE.
CC      -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC      -1- SIMILARITY: SOME SIMILARITIES EXIST BETWEEN EBNA 4, 5, AND 6.
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: V01555; CAA24859.1; -
CC      Pfam: PF05009; EBNA-3; 1.
CC      Nucleic protein; Repeat.
CC      FT      DOMAIN 551 610
CC      FT      DOMAIN 741 779
CC      POLY-ARG. 10 X 5 AA TANDEM REPEATS.
CC      3 X 13 AA TANDEM REPEATS.

```


SQ SEQUENCE 992 AA; 109129 MW; 39BBA9BC51BD84 CRC64;
 Query Match 13.6%; Score 160.5; DB 1; Length 992;
 Best Local Similarity 26.1%; Pred. No. 0.018;
 Matches 62; Conservative 22; Mismatches 83; Indels 71; Gaps 8;

QY 33 PLHPS-LNIPYIGIRNLPLPYLRPVNTVPSYSGN-----TYNDT 70
 Db 464 PAPHFVPEWPIVILHQPPEV-DKPVVVKETPPSPRRRGACVYDDVIEVDVETTES 522
 QY 71 GLPSPWILTSBPGFVYVHIRGFLATQNLVPLPPRGFPVPSRFSMAAPAPPIA 130
 Db 523 S-----SVSGPNKHKRGQDFORSGRQKAAAPTYS-----SDTGPVAGPPA 568
 QY 131 AEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAE-----APVAAEPAAEP-- 183
 Db 569 AGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPAAGPPA 628
 QY 184 -----VGVEPPA-----PEPPAEPATKAPAPPEHPSPL 214
 Db 629 VAMFEREQLPOSTGRKQCFCEWEMKRAGREITOMQDEPSSHLSQATQPTTPRSPMAPSV 686

RESULT 14
 ENAM_MOUSE
 ID ENAM_MOUSE STANDARD; PRT; 1274 AA.
 AC 055196;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Enamelin precursor.
 GN ENAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Enamel epithelium;
 RX MEDLINE=2051493; PubMed=11062988;
 RA Hu C.-C., Simmer J.P., Bartlett J.D., Nanci A., Qian Q., Zhang C.,
 RA Ryu O.H., Xue J., Fukae M., Uchida T., McDonald M.;
 RT "Murine enamel: cDNA and derived protein sequences."
 RL Connect. Tissue Res. 39:47-61(1998).
 CC -1- FUNCTION: Involved in the mineralization and structural
 CC organization of enamel. Involved in the extension of enamel during
 CC the secretory stage of dental enamel formation.
 CC -1- SUBCELLULAR LOCATION: Secreted. Extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Expressed in developing teeth.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U82698; AAB94312.1; -
 DR MGD; MGI:133772; Enam.
 DR PIR; T37193;
 DR GO; GO:0005578; C:extracellular matrix; ISS.
 DR GO; GO:0030345; F:structural constituent of tooth enamel; ISS.
 DR GO; GO:0030282; F:bone mineralization; ISS.
 DR GO; GO:0042476; P:odontogenesis; ISS.
 KM Biomineralization; Extracellular matrix; Glycoprotein; Signal.
 FT SIGNAL 38
 FT CHAIN 1
 FT CARBOHYD 39 1274
 FT CARBOHYD 130 130
 FT CARBOHYD 252 252
 FT CARBOHYD 259 259
 FT CARBOHYD 269 269
 FT CARBOHYD 300 300
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1066 1066
 SQ SEQUENCE 1274 AA; 140954 MW; F9DBD1CC9D327143 CRC64;
 Query Match 13.4%; Score 158; DB 1; Length 1274;
 Best Local Similarity 28.0%; Pred. No. 0.03;
 Matches 75; Conservative 18; Mismatches 95; Indels 80; Gaps 16;

QY 17 KRRPFIEGDNDGDHPLHPSLN-----IPVG-----I 44
 Db 220 KEKOPKEDEPPDPPEEASTNSTVPDANATQSTPEGNDTSPICGTGPGNACNPTV 289
 QY 45 RN-LPPLLY-----PEVNTVPSYSG--NTYTDGLPSYP-----WILT-----SFG 83
 Db 220 QNGVPPPKVAVSGGVKSGQIPWRPSGPNYENPYNYISERQMTTGQPRQNGG 349
 QY 84 PYYVYHNG-----PELATQNLVPLPPRGFPV--PSPRFSMAAPAP 128
 Db 350 YRNQVERGPQWNSPFWMEKQATRGNPTYGPEPTSGVYVAGNPVHFGRLPGRNKEF 409
 QY 129 IAAEPAAAPLT-ATP-----VAAEPAGAP-VAAEPAAEP-VGAEPAAEP-VAAEP 178
 Db 410 VGANPASNKKPVGANPASNKKPFVGANPASNKKPFVGANPASNKKPFVGANPASNKKPVGANP 469
 QY 179 AAAP-VGVEPPAAEPSP-PAEPATKPA 204
 Db 470 ASNKEPFGANPANKPSIGTNPAANKPS 497

RESULT 15
 CCPA_ACEXY
 ID CCPA_ACEXY STANDARD; PRT; 353 AA.
 AC P37697;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cellulose complementing protein.
 GN CCPAX.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Acetobacteraceae; Gluconacetobacter.
 OX NCBI_TaxID=28448;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 23769;
 RX MEDLINE=94131945; PubMed=8300521;
 RA Standaal R., Iversen T.-G., Coucheron D.H., Fjaervik E., Blatny J.M.,
 RA Valla S.;
 RT "A new gene required for cellulose production and a gene encoding
 RT cellulytic activity in Acetobacter xylinum are colocalized with the
 RT bcs operon."
 RT J. Bacteriol. 176:665-672(1994).
 RN [2]
 RP SEQUENCE OF 232-353 FROM N.A.
 RC STRAIN=ATCC 53582;
 RX MEDLINE=91322509; PubMed=1830823;
 RA Saxena I.M., Lin P.C., Brown R.M. Jr.;
 RT "Identification of a new gene in an operon for cellulose biosynthesis
 RT in Acetobacter xylinum."
 RT Plant Mol. Biol. 16:947-954(1991).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M96060; AAA16970.1; -
 DR EMBL; X54676; -; NOT ANNOTATED_CDS.
 DR PIR; B36963; B36963.
 DR VARIANT 232 232
 FT VARIANT 244 250
 FT R -> Q (IN STRAIN ATCC 53582).
 FT ARLSPRE -> HASAAG (IN STRAIN ATCC 53582).

Query Match	13.1%	Score 154;	DB 1;	Length 353;
Best Local Similarity	29.9%;	Pred. No. 0.017;		
Matches	52;	Conservative	9;	Mismatches 73;
				Indels 40;
				Gaps 7;

```

Db      37 YRP-----FVDRSPDVTGVE-----AVERHFDQAEH-----DTAVEEQVTPAP----- 75

```

```
Db      76 -----QIAVAPRPVDPRAIVTETAPRPV--VVSAPVTYEPRAAAVPAEPVQEA 127
```

Db 128 PVQAAPVPPAP--VPPIAEQAAPPAPDPASVPPYANVAAPVPPDPAPVTPAPQA 179

Search completed: January 15, 2004, 15:12:28
Job time : 19 secs

A:Molecule type: DNA
A:Residues: 1-395 <WHI>
A:Cross-references: GB:AE001946; GB:AE000513; NID:g6458655; PIDN:AAFI0516.1; PID:g645866
A:Experimental source: strain RI
C:Genetics:
A:Gene: DR0938
A:Map position: 1

Query Match 15.8%; Score 186.5; DB 2; Length 395;
Best Local Similarity 35.0%; Pred. No. 0.0001;
Matches 57; Conservative 11; Mismatches 64; Indels 31; Gaps 9;

QY 60 PSYF-GNTYDTGLSYFPIILTSFGSPYVYHIGFPLATQLWVPLPRGPFVPSRPF 118
DB 53 PAEPVGAQPTTSTIVSFGAVVADSPQT---AAPAQATTAGKIPPA--PAAP----- 102
QY 119 SAAAAPAAP-----IAEPAAAPLTATPVAEPAAAGAPVAEPAAEPVGAEPAA 170
DB 103 ----APKIPPLPRPRMPPAPPTPRPVEETTTAASPTQPPVQAAPATQPTPTQTAA 158
QY 171 EAPVAAEPAAEPVGEVPA--EEGSPA--EPAT-AKPAEP 208
DB 159 QAPATQPPATPAF---EPAAAPAPATTPPEPATPEPAAPAP 198

RESULT 7

Hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A:Accession: E66255
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nasser, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E66255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <STO>
A:Cross-references: GB:AE005172; NID:g3157926; PIDN:AACI7609.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 15.5%; Score 183; DB 2; Length 744;
Best Local Similarity 28.9%; Pred. No. 0.00031;
Matches 55; Conservative 14; Mismatches 93; Indels 28; Gaps 6;

QY 33 PLHPLNIPYGRINLPPLIYRPVNTVPSYNGNTYTDGLSPY-----PILTSRGP 85
DB 461 PPSPPPPPPYVSSPPPPYV---SSPPPPYVSSPPPPYVSSPPPPYVSSPPPP 516
QY 86 YUHHRGFPLATQLWVPLPRGPFVPSRFFSAAAAPAA--PPIAEPAAALITATP 143
DB 517 YVY-----SSPPPP---PSPPPCPSSPPPPVYVAVTQSPPPSPSVYYP 562
QY 144 VAAEPAAAPVAEPAAEPVGAEPAAEPVAAEPAAEPVGEVPAEPSPSAEPATAP 203
DB 563 VTQSPPPSPSVYVPPVNTSPPPSPSVYVPPVTSPPSPSVYVPPVTSPPSPSVYYP 622
QY 204 AAPRPPSPS 213
DB 623 VTPSP-PPPS 631

RESULT 8

AA0215
C:Species: human herpesvirus 4 (strain B95-8)
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999

TCD antigen - Trypanosoma cruzi

C:Species: Trypanosoma cruzi
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
A:Accession: AA0215
R:Burns Jr., J.M.; Shreffler, W.G.; Roman, D.E.; Sleath, P.R.; March, C.J.; Reed, S.G.
Proc. Natl. Acad. Sci. U.S.A. 89, 1239-1243, 1992
A:Title: Identification and synthesis of a major conserved antigenic epitope of Trypanos
A:Reference number: AA0215; MUID:92159024; PMID:1371355
A:Accession: AA0215
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <BUR>
A:Note: sequence extracted from NCBI backbone (NCBI:82113, NCBI:82117)
C:Superfamily: histone H1

Query Match 15.2%; Score 179; DB 2; Length 207;
Best Local Similarity 41.0%; Pred. No. 0.00016;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAAPAPPIAEPAAAPLTATPVAEPAAAGAPVAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAPPPAEPKSAEPKPAEPKSGPPAEPKSAEPKPAEPK 63
QY 174 VAAEPAAEPVGEVPAEPSP-----APATAPPAEPSPSPS 213
DB 64 KPAPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPK 108

RESULT 9

Translational initiation factor IF-2 - Stigmatella aurantiaca
C:Species: Stigmatella aurantiaca
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Feb-2001
A:Accession: T43226
R:Bremund, L.; Laalami, S.; Derijard, B.; Cennatempo, Y.
J. Bacteriol. 179, 2348-2355, 1997
A:Title: Translational initiation factor IF2 of the myxobacterium Stigmatella aurantiaca:
A:Reference number: Z22352; MUID:97234648; PMID:9079922
A:Accession: T43226
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1054 <BRE>
A:Cross-references: EMBL:X87940; PIDN:CAA61162.1
A:Experimental source: strain DM4
C:Genetics:
A:Gene: infB
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homc
C:Keywords: nucleotide binding; P-loop; protein biosynthesis
F:555-664/Domain: translation elongation factor Tu homology <ETU>
F:561-568/Region: nucleotide-binding motif A (P-loop)

Query Match 14.9%; Score 175; DB 2; Length 1054;
Best Local Similarity 38.3%; Pred. No. 0.0013;
Matches 49; Conservative 8; Mismatches 51; Indels 20; Gaps 4;

QY 96 ATQNVPLPRGPFVPSRFFSAAAAPAPPIA--EPAAAAPLTAT--PVAEPAAAG 151
DB 98 ASDVSSPSPSP-----VHEASGAEEAASERVEAAAVQBPVAEAPRAAASEPAAAPKAT 152
QY 152 APVAEPAAEPVGAEPAAEPVAAEPAAEPVGEVPAEPSP-----APAT 200
DB 153 APVAEPVVEAPKAAAPVAEPVTEAPKTEAPVAAPVIAEAPTPAPARTPEVPTSGRAAS 212
QY 201 AKPAEP 208
DB 213 CRGAAPLP 220

RESULT 10

QOBB8
BFLI protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999


```

Db      16 GVDSDTAVNMDSPT-AADGPIEVSAGAPVSEHAKRPLERQALFNTGSPISSTTAAEA 74
Qy      71 GLPSTPWLITSPGFPYVTHIRGFLATQ-----LVNPPLEPRGFPVPP 114
Db      75 KVPSTLEREGSGTQGETVHIKPAVPAESGTDSSKADPSATVAVLIQISPEEVGVPMTPT 134
Qy      115 SRFFSAAAAPAPPIAAPP-AAAAPLTATPVAA-----PPAAGAPVAAPPAAPVGAAP 169
Db      135 D-LPPASDAGPDVRAEPDGGTAP--ATPAESDNREPPAAA--AAEPAAEP--AAEPA 186
Qy      170 AE--APVAAPAAEA-----PVGVEPAAPESPAPAPATAPAPAPHPSP 212
Db      187 AEPAAPPAAPPAAPPAAPVPTAEASAGAVPTQEEPAAPAAASATPAEPAPAAAP 239

```

RESULT 14

G87675
Arylsterase-related protein (imported) - Caulobacter crescentus

C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: G87675

R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: G87675

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <STO>

A:Cross-references: GB:AE005673; NID:g13425157; PIDN:AAK25403.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3441

Query Match 14.6%; Score 171.5; DB 2; Length 438;

Best Local Similarity 33.3%; Pred. No. 0.00094;

Matches 68; Conservative 14; Mismatches 59; Indels 63; Gaps 11;

```

Qy      64 GNTYDTGTPSY-----WLTSPGF-----PYVHIRGFPPLATQLVNPPLEPRGFPVPP 114
Db      219 GRKVASADIKIYNNNAHVIDEPNTMKLIDVIAWLKGTQAPAKVBPAP----- 270
Qy      115 SRFFSAAAAPAPPIAAPPAAAPLTATPV-AAEPAA-----GAPVAAEPA-----A 160
Db      271 ----PAAAPAPAPAPAKAPEPAAAP---EPVKAAPASPAPKAKAPKAPKADAKPKATA 323
Qy      161 EAPVG-----APPAEAPV--AAEPAAEPVGVPEPAEESPAPF----- 197
Db      324 KAPVAKKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 383
Qy      198 --PATAPAPAPPEHPSPSLEQANQ 219
Db      384 TAPAKKAPAPKAPKAPKAPKAPKAPKATATK 407

```

RESULT 15

T17547

Proline-rich protein A57R - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17547

R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: T17547

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-544 <GRA>

A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAK6425.1

A:Experimental source: specific host Chlorella strain NC64A

A:Gene: A57R

Query Match 14.5%; Score 171; DB 2; Length 544;

Best Local Similarity 27.8%; Pred. No. 0.0012;

Matches 58; Conservative 19; Mismatches 106; Indels 26; Gaps 4;

```

Qy      2 KLLIMACIVCAFAFKRRF---PFIQEDNDGHPHPSLNIPIYGRILPPLYYRPVNT 58
Db      4 RVIFFLITVTTAFAPPEPFTNQIVTTQTVVNSFKPPSPKPPSPKPPSPKPPSP 63
Qy      59 VPSYSGNTYDTGTPSYFWLTSPGFPYVTHIRGFLATQLVNPPLEPRGFPVPPSRFP 118
Db      64 KPSPKPPSPKPPSPKPP-----SP-----KPPSPKPPSP-KPPSPKPP 100
Qy      119 SAAAAPAAPPIAAPPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEP 178
Db      101 PSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPSPKPPAPPEPAPPEPAPPEP 160
Qy      179 AAAPVGVPEPAAPESPAPAPATAPAPAP 207
Db      161 PAPPEPAPPEPAPPEPAPPEPAPPEPAPPE 189

```

Search completed: January 15, 2004, 15:14:44
Job time : 23 secs

THIS PAGE BLANK (USPTO)

Qy	Db	Qy	Db	Qy	Db	Qy	Db	Qy	Db
1	1	61	61	121	121	181	181	241	241
MLLLVACIVCAFAARKRPFPIGDDDDGHHLPISLNIPIYGRINLPPLYYRPVNTVP	MLLLVACIVCAFAARKRPFPIGDDDDGHHLPISLNIPIYGRINLPPLYYRPVNTVP	SYGNTYTTGTGLSPFWILITSGFPYYVHHIRGFPLATOLNPPLPRGFPVPPSPRFFSA	SYGNTYTTGTGLSPFWILITSGFPYYVHHIRGFPLATOLNPPLPRGFPVPPSPRFFSA	AAAPAAPPIAAEPPAAAAPLTATPVAAEPAAGAPVAAEPAEAPVGAEEPAEAPVAAEPAA	AAAPAAPPIAAEPPAAAAPLTATPVAAEPAAGAPVAAEPAEAPVGAEEPAEAPVAAEPAA	BAFVGVEPAABEBSPAEPATKAPAAPBHPSPSELANQ	BAFVGVEPAABEBSPAEPATKAPAAPBHPSPSELANQ	AAAPAAPPIAAEPPAAAAPLTATPVAAEPAAGAPVAAEPAEAPVGAEEPAEAPVAAEPAA	AAAPAAPPIAAEPPAAAAPLTATPVAAEPAAGAPVAAEPAEAPVGAEEPAEAPVAAEPAA
60	60	120	120	180	180	240	240	300	300

Db 61 SYPGNTYDTGLPSYFWILTSRGPYVYHINRFPPLATQUNVPLPRGPFPPPSRFFSA 120
 Qy 121 AAAPAPPIAAEPAAAAELTATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 Db 121 AAAPAPPIAAEPAAAAELTATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 Qy 181 EAPVGEPAEPEEPSPAEPAATKAPAPBPSPSLEQANQ 219
 Db 181 EAPVGEPAEPEEPSPAEPAATKAPAPBPSPSLEQANQ 219

RESULT 5

US-10-231-417-192
 ; Sequence 192, Application US/10231417
 ; Publication No. US20030176681A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Feng et al.
 ; TITLE OF INVENTION: 148 Human Secreted Proteins
 ; FILE REFERENCE: P2019P1
 ; CURRENT APPLICATION NUMBER: US/10/231,417
 ; CURRENT FILING DATE: 2002-08-30
 ; PRIOR APPLICATION NUMBER: US/09/296,622
 ; PRIOR FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 619
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 192
 ; LENGTH: 221
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1159)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; LOCATION: (221)
 ; NAME/KEY: SITE
 ; LOCATION: (221)
 ; OTHER INFORMATION: Xaa equals stop translation
 ; US-10-231-417-192

Query Match 94.5%; Score 1113.5; DB 12; Length 221;
 Best Local Similarity 95.0%; Pred. No. 1.1e-72;
 Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

Qy 1 MGLIMACTICVAFARKRRFPETGDDNDGHLPSLNIPIYGINLPPPLYRPVNTVP 60
 Db 1 MGLIMACTICVAFARKRRFPETGDDNDGHLPSLNIPIYGINLPPPLYRPVNTVP 60
 Qy 61 SYPGNTYDTGLPSYFWILTSRGPYVYHINRFPPLATQUNVPLPRGPFPPPSRFFSA 120
 Db 61 SYPGNTYDTGLPSYFWILTSRGPYVYHINRFPPLATQUNVPLPRGPFPPPSRFFSA 120
 Qy 121 AAAPAPPIAAEPAAAAELTATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 179
 Db 121 AAAPAPPIAAEPAAAAELTATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 Qy 180 AEPVGEPAEPEEPSPAEPAATKAPAPBPSPSLEQANQ 219
 Db 181 AEPVGEPAEPEEPSPAEPAATKAPAPBPSPSLEQANQ 220

RESULT 6

US-10-316-253-95
 ; Sequence 95, Application US/10316253
 ; Publication No. US20030162706A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Procter & Gamble Company
 ; APPLICANT: Peters, Kevin
 ; APPLICANT: Thompson, Larry
 ; APPLICANT: Wang, Feng
 ; APPLICANT: Greis, Kenneth
 ; TITLE OF INVENTION: Angiogenesis Modulating Proteins
 ; FILE REFERENCE: 8865M

; CURRENT APPLICATION NUMBER: US/10/316,253
 ; CURRENT FILING DATE: 2002-12-10
 ; PRIOR APPLICATION NUMBER: US 60/355,295
 ; PRIOR FILING DATE: 2002-02-08
 ; NUMBER OF SEQ ID NOS: 308
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 95
 ; LENGTH: 2657
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-10-316-253-95

Query Match 15.4%; Score 181; DB 12; Length 2657;
 Best Local Similarity 29.1%; Pred. No. 0.00012;
 Matches 57; Conservative 22; Mismatches 61; Indels 56; Gaps 8;

Qy 40 IPYGINLPPPP-----LYRPVNTVPSYPGNTYDTGLPSYFWILTSRGPYVYHINRFP 93
 Db 2239 VKFGHKQINPHTANSLSITSKVTTIK--PANSSEKPTAI-----VNLTP----- 2280
 Qy 94 PLATQUNVPLPRGPFPPV-----PSRFFSAAAAPAP-----PIAEP 133
 Db 2281 -----AKPAPAPPAAPQPVLAKEPPAQAPAPAPKAPASAKLVPPQPVHVQAPAPQT 2333
 Qy 134 AAAPLTATPVAAEPAGAPVAAEPAAEA-PVGAEPAAEPVAAEPAAEPVGAEPAAEP 192
 Db 2334 ASVREPAKAPAPPOPAKAPVPAKAPVAPQAPPAKAPVPAKAPVPAKAPVPAKAPVPA 2389
 Qy 193 PSPAPAPAPAPAP 208
 Db 2390 PMPAPVLTGSAAVXP 2405

RESULT 7

US-10-264-049-2409
 ; Sequence 2409, Application US/10264049
 ; Publication No. US20040005579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Biree et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA133P1
 ; CURRENT APPLICATION NUMBER: US/10/264,049
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: US 60/209,467
 ; PRIOR FILING DATE: 2000-06-07
 ; NUMBER OF SEQ ID NOS: 4360
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 2409
 ; LENGTH: 903
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (110)
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (15)
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (16)
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (795)
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 ; US-10-264-049-2409

Query Match 14.9%; Score 175.5; DB 12; Length 903;
 Best Local Similarity 32.3%; Pred. No. 9.9e-05;

;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-023-523-43

Query Match 14.7%; Score 173; DB 14; Length 538;
Best Local Similarity 36.8%; Pred. No. 8.9e-05;
Matches 49; Conservative 7; Mismatches 47; Indels 30; Gaps 5;

QY 102 PPLPRGPFPPSPFSSAAAPPAAPPIAAEPPAAAPLTATVAAEPAAAGVAAEPAAE 161
DB 111 PPRARPGAP-----AAAAAAPPPTPAPPPPPAPVAAAAPAPAAAAAATAAPSP 162
QY 162 APVGAEPAAE-APVAA-EPAEAPVGVPE-----AABEP-----SPABA 199
DB 163 GBAQGPRAQRAAPPAAPPAPVAPVAPGPRAPPAVAAAPLPPPPQPPAPQ 222
QY 200 TAKPAAPEPHPS 212
DB 223 QQQPPPPQPP 235

RESULT 11

US-10-029-386-33686
; Sequence 33686, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Seqmax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 33686
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALJ37853.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.84
; OTHER INFORMATION: SWISSPROT HIT: Q02910, EVALUE 2.00e-03
US-10-029-386-33686

Query Match 14.7%; Score 173; DB 12; Length 980;
Best Local Similarity 31.1%; Pred. No. 0.00016;
Matches 68; Conservative 13; Mismatches 112; Indels 26; Gaps 9;

QY 6 WACIVCAFAARRRRPFIGEDNDGHLPLPSLNT-----PYGIRLPPPLYRPPVNT 58
DB 531 WAAALATV-----PITEDEGTPEG-PVTPTATVAPPEPDPAVAVSTPEEPASPA 582
QY 59 VPSYGNITTDGLSGYPILTSFGPPYHYHNGPPLATQLANVPLPPGPFVPP-PSRF 117
DB 563 VTPPEPTSPAAAVPT-PEEPTSPAAAVPPPEPTSPAAAVPTPEEPTSPAAAVPTPEEP 641
QY 118 PFAAABAPAPPIAAEPPAAAPLTATVAAEPAAAGVAAE-PAEAPVGAEPAAEAPV 174
DB 642 TSPAAAVPTPEEPTSPAAAVPTPEEPTSPAAAVPTPEEPTSPAAAVPTPEEPA-SRA 697
QY 175 AAEPAEAPVGAEPAAEPPSPAEPAETAKPAEPHPS 213
DB 698 AAVPTPEEP--ASPAAVPTPEEPAPPAVAVPTPEESAS 724

RESULT 12
US-10-108-260A-3849
; Sequence 3849, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE

;; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA

;; FILE REFERENCE: HI-A0106
;; CURRENT APPLICATION NUMBER: US/10/108,260A
;; CURRENT FILING DATE: 2002-03-27

;; NUMBER OF SEQ ID NOS: 5458
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO: 3849

;; LENGTH: 693
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-108-260A-3849

Query Match 14.6%; Score 171.5; DB 12; Length 693;
Best Local Similarity 31.6%; Pred. No. 0.00015;
Matches 62; Conservative 17; Mismatches 78; Indels 39; Gaps 11;

QY 33 PLHPSLNPYGIRLPPPLYRPPVNTVPSYGNITTDGLSPYILTSFGPPYHYHNG 92
DB 253 PVPSPATPP---SQAPFSLAAPPLQVPPSPASP-----PMSPATPPPPQAP----- 297
QY 93 PPLAT-QANVPLPPGPFVPPSPFSSAAAP-----AAPPIAAEPPAAAPLTATPVA 145
DB 298 -PLAAPPLQVPPSP 351
QY 146 AEPAGAPVAAEPAAEAPVGAEPAAE-PAVAAEPAAEAPVGAEPAAEPPSPSPSPSP 202
DB 352 KPFPQAPPALATPPLQA-LPSBPSPFGQAPPS--PSASLP--MSPLATPPQAPPVLA 406
QY 203 P---AAEPHPSPSLE 215
DB 407 PLQVPPSP 422

RESULT 13

US-10-084-843-142
; Sequence 142, Application US/10084843
; Publication No. US20030143243A1

; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

; Skelky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Iwardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.

;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

;; NUMBER OF SEQUENCES: 355

;; CORRESPONDENCE ADDRESS:

;; ADDRESS: SEED and BERRY LLP

;; STREET: 6300 Columbia Center, 701 Fifth Avenue

;; CITY: Seattle

;; STATE: Washington

;; COUNTRY: USA

;; ZIP: 98104-7092

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent In Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/10/084,843

;; FILING DATE: 25-Feb-2002

;; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US/09/072,967

;; FILING DATE: 05-MAY-1998

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Makl, David J.

;; REGISTRATION NUMBER: 31,392

;; REFERENCE/DOCKET NUMBER: 210121.411C9

Db 348 A--PAPETKSDAPEPAAAAADPAPETKSDAAAAADPAPGTKADAA 387

Search completed: January 15, 2004, 15:19:47
Job time : 34 secs

THIS PAGE BLANK (USPTO)

1 TITLE OF INVENTION: Against T. cruzi Infection

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSER: Immunex Corporation
STREET: 51 University street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/836,642
FILING DATE: 14-FEB-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: REED
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-836-642-2

Query Match          15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAPAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPK 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPDPSPS 213
DB 64 KPAPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPK 108

RESULT 3
US-08-169-563-2
Sequence 2, Application US/08169563
Patent No. 5413912
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
TITLE OF INVENTION: Against T. cruzi Infection
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSER: Immunex Corporation
STREET: 51 University street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.1
SOFTWARE: Microsoft Word For Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,563
FILING DATE: 17-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/836,642
FILING DATE: 14-FEB-1992
CLASSIFICATION: 435
```

```

ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: REED-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-563-2

Query Match          15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAPAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPK 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPDPSPS 213
DB 64 KPAPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPK 108

RESULT 4
US-08-403-379A-6
Sequence 6, Application US/08403379A
Patent No. 5756662
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,379A
FILING DATE: 14-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Naki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-379A-6

Query Match          15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAPAPPIAEPAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEP 173
```

Db 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
Qy 174 VAAPPAAPVGVPEPAEPPSP-----AEPATAPAEPEPPSPS 213
Db 64 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 5
US-08-557-309B-50
; Sequence 50, Application US/08557309B
; Patent No. 5916572

GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557.309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 210121.422

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-08-557-309B-50

Query Match 15.2%; Score 179; DB 2; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

Qy 114 PSRFSAAAAPAPPIAAPELTTPTVVAABPAAGAPVVAABPAAPGAEPPAAP 173

Db 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63

Qy 174 VAAPPAAPVGVPEPAEPPSP-----AEPATAPAEPEPPSPS 213

Db 64 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 6
US-08-929-414-6

; Sequence 6, Application US/08929414

; Patent No. 5942403

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Houghton, Raymond

; APPLICANT: Skeiky, Yasir A.W.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION

; TITLE OF INVENTION: OF T. CRUZI INFECTION

; NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/929.414

FILING DATE: 15-SEP-1997

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.406C1

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 207 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-929-414-6

Query Match 15.2%; Score 179; DB 2; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

Qy 114 PSRFSAAAAPAPPIAAPELTTPTVVAABPAAGAPVVAABPAAPGAEPPAAP 173

Db 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63

Qy 174 VAAPPAAPVGVPEPAEPPSP-----AEPATAPAEPEPPSPS 213

Db 64 KPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 7

US-08-834-306-50

; Sequence 50, Application US/08834306

; Patent No. 6054135

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

; NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/834.306

FILING DATE: 15-APR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-50

Query Match 15.2%; Score 179; DB 3; Length 207;
Best Local Similarity 41.0%; Pred. No. 4,1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PRRFSAAPAPAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATAKPAEPHPSPS 213
DB 64 KPAPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 8
US-08-993-674A-50
Sequence 50, Application US/08993674A
Patent No. 6228372

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-50

Query Match 15.2%; Score 179; DB 3; Length 207;
Best Local Similarity 41.0%; Pred. No. 4,1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PRRFSAAPAPAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATAKPAEPHPSPS 213
DB 64 KPAPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 9
US-09-256-976-50
Sequence 50, Application US/09256976
Patent No. 6419933

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
FILE REFERENCE: 210121.422C3
CURRENT APPLICATION NUMBER: US/09/256,976
CURRENT FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 50
LENGTH: 207
TYPE: PR
ORGANISM: Trypanosoma cruzi
US-09-256-976-50

Query Match 15.2%; Score 179; DB 4; Length 207;
Best Local Similarity 41.0%; Pred. No. 4,1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PRRFSAAPAPAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATAKPAEPHPSPS 213
DB 64 KPAPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAES 108

RESULT 10
US-08-818-112-142

Sequence 142, Application US/08818112
Patent No. 6280969

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.

APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vardick, Thomas S.
APPLICANT: Wardzik, Daniel R.
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:29:52 ; Search time 33 Seconds
(without alignments)
1357.012 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MLLLMACIVCAFAFKRRF.....TAKPAPRPHSPSLEQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 762491 seqs, 204481190 residues

Word size : 0

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	100.0	219	9	US-09-922-480-2
2	219	100.0	219	9	US-09-923-236-2
3	219	100.0	219	10	US-09-923-469-2
4	150	68.5	221	12	US-10-231-417-192
5	140	63.9	219	11	US-09-746-783-106
6	9	4.1	798	15	US-10-156-761-13162
7	9	4.1	2969	10	US-09-738-626-4434
8	3.7	1332	12	US-10-289-762-96	
9	3.7	206	15	US-10-219-220-250	
10	3.7	249	15	US-10-180-375-82	
11	3.7	342	12	US-10-369-493-12383	
12	3.7	367	12	US-10-369-493-7317	
13	3.7	383	10	US-09-788-345-10	
14	3.7	383	12	US-10-337-312-10	
15	3.7	412	10	US-09-788-345-12	
					Sequence 2, Appl1
					Sequence 2, Appl1
					Sequence 192, Appl1
					Sequence 106, Appl1
					Sequence 13161, Appl1
					Sequence 4434, Appl1
					Sequence 96, Appl1
					Sequence 250, Appl1
					Sequence 82, Appl1
					Sequence 12383, Appl1
					Sequence 7317, Appl1
					Sequence 10, Appl1
					Sequence 12, Appl1

16	8	3.7	412	12	US-10-337-312-12	Sequence 12, Appl
17	8	3.7	428	10	US-09-906-514-4	Sequence 4, Appl
18	8	3.7	428	12	US-10-177-478-14	Sequence 14, Appl
19	8	3.7	454	15	US-10-156-761-13939	Sequence 13939, A
20	8	3.7	511	9	US-09-864-761-34590	Sequence 34590, A
21	8	3.7	695	12	US-10-029-366-32280	Sequence 32280, A
22	8	3.7	695	12	US-10-369-493-15364	Sequence 15364, A
23	8	3.7	850	12	US-10-369-493-10111	Sequence 10111, A
24	8	3.7	4307	12	US-10-369-493-5698	Sequence 5698, Ap
25	8	3.7	4307	12	US-10-369-493-5699	Sequence 5699, Ap
26	8	3.7	4307	12	US-10-369-493-5700	Sequence 5700, Ap
27	8	3.7	8026	12	US-10-132-124-12	Sequence 12, Appl
28	8	3.7	16	12	US-10-397-551-43	Sequence 43, Appl
29	7	3.2	21	14	US-10-124-800-15	Sequence 15, Appl
30	7	3.2	27	11	US-09-974-879-106	Sequence 306, App
31	7	3.2	27	11	US-09-305-726-306	Sequence 306, App
32	7	3.2	27	12	US-09-818-663-306	Sequence 306, App
33	7	3.2	48	9	US-09-739-907-177	Sequence 177, App
34	7	3.2	48	12	US-09-938-671-177	Sequence 177, App
35	7	3.2	60	9	US-09-864-761-33554	Sequence 33554, A
36	7	3.2	60	9	US-09-864-761-33624	Sequence 33624, A
37	7	3.2	60	9	US-09-864-761-34127	Sequence 34127, A
38	7	3.2	61	9	US-09-864-761-35187	Sequence 35187, A
39	7	3.2	65	9	US-09-864-761-36194	Sequence 36194, A
40	7	3.2	78	11	US-09-820-843A-89	Sequence 89, Appl
41	7	3.2	88	9	US-09-764-853-503	Sequence 503, App
42	7	3.2	91	9	US-09-867-550-738	Sequence 738, App
43	7	3.2	95	9	US-09-739-907-178	Sequence 178, App
44	7	3.2	95	12	US-09-938-671-178	Sequence 178, App
45	7	3.2	105	12	US-10-104-047-2874	Sequence 2874, Ap

ALIGNMENTS

```
RESULT 1
US-09-922-480-2
; Sequence 2, Application US/09922480
; Patent No. US20020081701A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZS163 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-480-2
```

Query Match	100.0%	Score 219;	DB 9;	Length 219;
Best Local Similarity	100.0%	Pred. No. 5.6e-185;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLLLMACIVCAFAFKRRFPPIGDDDDGRLHPSINIPGINLPPPLYYRVNTYP	60	
DB	1	MLLLMACIVCAFAFKRRFPPIGDDDDGRLHPSINIPGINLPPPLYYRVNTYP	60	
QY	61	SYNGATYTDGSPYFWILTSRGPYYVHINRFPPLATOLNVPPLPPRGPFVPSRFFSA	120	
DB	61	SYNGATYTDGSPYFWILTSRGPYYVHINRFPPLATOLNVPPLPPRGPFVPSRFFSA	120	
QY	121	AAAPAPPIAAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA	180	
DB	121	AAAPAPPIAAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEPVAAEPAA	180	
QY	161	EAPVGEPAEPAAEPAAEPATATKAPAPRPHSPSLEQANQ	219	

Db 181 EAPVGEPAAEPSPAEPATAKPAAPDEPHPSLSLEQANQ 219

RESULT 2
US-09-923-236-2
Sequence 2, Application US/09923236

Patent No. US20020090677A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/923,236
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-923-236-2

Query Match Best Local Similarity 100.0%; Score 219; DB 9; Length 219;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACIVCAFAARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPPLYRRPVNTVP 60

Db 1 MKLLMACIVCAFAARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPPLYRRPVNTVP 60

QY 61 SYPGNTYTDGTPSYWILTSFGFPYVYHIRGFPPLATOLNVPPLPRGFPVPSRFFSA 120

Db 61 SYPGNTYTDGTPSYWILTSFGFPYVYHIRGFPPLATOLNVPPLPRGFPVPSRFFSA 120

QY 121 AAPAAPPIAAEPAAAPLTAIPVAAPAAAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180

Db 121 AAPAAPPIAAEPAAAPLTAIPVAAPAAAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATAKPAAPDEPHPSLSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATAKPAAPDEPHPSLSLEQANQ 219

RESULT 3

US-09-922-469-2
Sequence 2, Application US/09922469

Patent No. US20020173027A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/922,469
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-469-2

Query Match Best Local Similarity 100.0%; Score 219; DB 10; Length 219;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACIVCAFAARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPPLYRRPVNTVP 60

Db 1 MKLLMACIVCAFAARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPPLYRRPVNTVP 60

QY 61 SYPGNTYTDGTPSYWILTSFGFPYVYHIRGFPPLATOLNVPPLPRGFPVPSRFFSA 120

Db 61 SYPGNTYTDGTPSYWILTSFGFPYVYHIRGFPPLATOLNVPPLPRGFPVPSRFFSA 120

QY 121 AAPAAPPIAAEPAAAPLTAIPVAAPAAAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180

Db 121 AAPAAPPIAAEPAAAPLTAIPVAAPAAAGAVAAEPAAEPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATAKPAAPDEPHPSLSLEQANQ 219

Db 181 EAPVGEPAAEPSPAEPATAKPAAPDEPHPSLSLEQANQ 219

RESULT 4
US-10-231-417-192
Sequence 192, Application US/10231417

Publication No. US20030176681A1

GENERAL INFORMATION:
APPLICANT: Feng et al.
TITLE OF INVENTION: 148 Human Secreted Proteins
FILE REFERENCE: P2019P1
CURRENT APPLICATION NUMBER: US/10/231,417
PRIOR FILING DATE: 2002-08-30
CURRENT APPLICATION NUMBER: US/09/296,622
PRIOR FILING DATE: 1999-04-23
NUMBER OF SEQ ID NOS: 619
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 192
LENGTH: 221
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: SITE
LOCATION: (159)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (221)
OTHER INFORMATION: Xaa equals stop translation
US-10-231-417-192

Query Match Best Local Similarity 100.0%; Score 150; DB 12; Length 221;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACIVCAFAARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPPLYRRPVNTVP 60

Db 1 MKLLMACIVCAFAARKRRFPFIGEDDNDGHLHPSLNIPYGINLPPPLYRRPVNTVP 60

QY 61 SYPGNTYTDGTPSYWILTSFGFPYVYHIRGFPPLATOLNVPPLPRGFPVPSRFFSA 120

Db 61 SYPGNTYTDGTPSYWILTSFGFPYVYHIRGFPPLATOLNVPPLPRGFPVPSRFFSA 120

QY 121 AAPAAPPIAAEPAAAPLTAIPVAAPAA 150

Db 121 AAPAAPPIAAEPAAAPLTAIPVAAPAA 150

RESULT 5
US-09-746-783-106

Sequence 106, Application US/09746783

Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth

McCoy, John M.
Lavallie, Edward R.
Racie, Lisa A.
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.
Howes, Steven H.

```

; Fuchel, Kim
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/746,783
; FILING DATE: 21-Dec-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Milasincic, Debra J.
; REGISTRATION NUMBER: 46,931
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-09-746-783-106

Query Match          63.9%; Score 140; DB 11; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.1e-115;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MGLLMACIVCAFAARKRFPFEGEDDNDGHPHPSLNIPIYGINLPPPLYRREVTVP 60
      1  MGLLMACIVCAFAARKRFPFEGEDDNDGHPHPSLNIPIYGINLPPPLYRREVTVP 60
DB      1  MGLLMACIVCAFAARKRFPFEGEDDNDGHPHPSLNIPIYGINLPPPLYRREVTVP 60

QY      61  SYPGNTYDTGTLPSYPMWITSGFPYVTHIRGEPPLATQLVNPLPPRGFPVPSRFPESA 120
      61  SYPGNTYDTGTLPSYPMWITSGFPYVTHIRGEPPLATQLVNPLPPRGFPVPSRFPESA 120
DB      61  SYPGNTYDTGTLPSYPMWITSGFPYVTHIRGEPPLATQLVNPLPPRGFPVPSRFPESA 120

QY      121  AAAPAPPAAPPAAPPAAPLT 140
      121  AAAPAPPAAPPAAPPAAPLT 140
DB      121  AAAPAPPAAPPAAPPAAPLT 140

RESULT 6
US-10-156-761-13162
; Sequence 13162, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

```

```

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13162
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13162

Query Match          4.1%; Score 9; DB 15; Length 798;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      198  PATAKPAAP 206
      198  PATAKPAAP 206
DB      29  PATAKPAAP 37

RESULT 7
US-09-738-626-4434
; Sequence 4434, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: YATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4434
; LENGTH: 2969
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4434

Query Match          4.1%; Score 9; DB 10; Length 2969;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      168  PAAPAPVAA 176
      168  PAAPAPVAA 176
DB      1663  PAAPAPVAA 1671

RESULT 8
US-10-289-762-96
; Sequence 96, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Giffaib, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 96
; LENGTH: 1332

```

TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-10-289-762-96

Query Match 3.7%; Score 8; DB 12; Length 132;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAEP 178
DB 53 EAPVAAEP 60

RESULT 9
US-10-219-220-250
Sequence 250, Application US/10219220
Publication No. US20030082724A1
GENERAL INFORMATION:

APPLICANT: Plim, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: Compositions affecting programmed cell
FILE REFERENCES: 11000.1022c1
CURRENT APPLICATION NUMBER: US/10/219,220
CURRENT FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
PRIOR FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 290
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 250
LENGTH: 206

TYPE: PRT
ORGANISM: EucaIyptus grandis
US-10-219-220-250

Query Match 3.7%; Score 8; DB 15; Length 206;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 112 AAAAPAP 119

RESULT 10
US-10-180-375-82
Sequence 82, Application US/10180375
Publication No. US20030126638A1
GENERAL INFORMATION:

APPLICANT: Allen, William B.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Famodu, OmoIayo O.
APPLICANT: Harwell, Leslie T.
APPLICANT: Helentjaris, Timothy
APPLICANT: Li, Changjiang
APPLICANT: Lowe, Keith
APPLICANT: Oliveira, Igor Cunha
APPLICANT: Shen, Bo

APPLICANT: Tarczyński, Mitchell C.
TITLE OF INVENTION: Alteration Of Oil Traits In Plants
FILE REFERENCE: BBI458 US NAI
CURRENT APPLICATION NUMBER: US/10/180,375
CURRENT FILING DATE: 2002-06-26

NUMBER OF SEQ ID NOS: 222
SOFTWARE: Microsoft Office 97
SEQ ID NO 82

LENGTH: 249
TYPE: PRT

ORGANISM: Trilicium aestivum
US-10-180-375-82

Query Match 3.7%; Score 8; DB 15; Length 249;
Best Local Similarity 100.0%; Pred. No. 42;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 VEPAAEP 193
DB 92 VEPAAEP 99

RESULT 11
US-10-369-493-12383
Sequence 12383, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12383
LENGTH: 342

TYPE: PRT
ORGANISM: Aspergillus nidulans

FEATURE:
NAME/KEY: unsure
LOCATION: (1)-(342)

OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12383

Query Match 3.7%; Score 8; DB 12; Length 342;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 126
DB 297 SAAAPAA 304

RESULT 12
US-10-369-493-7317
Sequence 7317, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 7317
LENGTH: 367

TYPE: PRT
ORGANISM: Burkholderia cepacia

US-10-369-493-7317

Query Match 3.7%; Score 8; DB 12; Length 367;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156

DB 71 AAGAPVAA 78

RESULT 13
US-09-788-345-10
; Sequence 10, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUEENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-09-788-345-10

Query Match 3.7%; Score 8; DB 10; Length 383;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 354 AAAAPAP 361

RESULT 14
US-10-337-312-10
; Sequence 10, Application US/10337312
; Publication No. US20030138451A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUEENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/10/337,312
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US/09/788,345
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-10-337-312-10

Query Match 3.7%; Score 8; DB 12; Length 383;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 354 AAAAPAP 361

DB 354 AAAAPAP 361

RESULT 15
US-09-788-345-12
; Sequence 12, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUEENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant sequence from vectors pQ31 and pMal
US-09-788-345-12

Query Match 3.7%; Score 8; DB 10; Length 412;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 383 AAAAPAP 390

Search completed: January 15, 2004, 15:35:27
Job time : 33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:28:37 ; Search time 21 Seconds
(without alignments)
441.242 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLIMACTIVCAFAKRRF.....TRKPAEPHPSLEQANO 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	US-09-527-345-2	Sequence 2, Appl.
2	9	4.1	322	US-08-428-414A-2	Sequence 2, Appl.
3	8	3.7	132	US-09-198-452A-96	Sequence 96, Appl.
4	8	3.7	168	US-09-252-991A-18199	Sequence 18199, A
5	8	3.7	190	US-09-252-991A-17963	Sequence 17963, A
6	8	3.7	307	US-09-252-991A-21588	Sequence 21588, A
7	8	3.7	383	US-09-471-396-3	Sequence 3, Appl.
8	8	3.7	383	US-09-788-345-10	Sequence 10, Appl.
9	8	3.7	410	US-09-252-991A-30606	Sequence 30606, A
10	8	3.7	412	US-09-471-396-1	Sequence 1, Appl.
11	8	3.7	412	US-09-788-345-12	Sequence 12, Appl.
12	8	3.7	549	US-09-252-991A-20196	Sequence 20196, A
13	8	3.7	575	US-08-653-648A-5	Sequence 5, Appl.
14	8	3.7	778	US-09-252-991A-37591	Sequence 27591, A
15	8	3.7	4545	US-08-804-227C-14	Sequence 14, Appl.
16	8	3.7	4550	US-08-804-227C-8	Sequence 8, Appl.
17	8	3.7	4550	US-08-804-198-2	Sequence 2, Appl.
18	8	3.2	31	US-09-248-588-27	Sequence 27, Appl.
19	7	3.2	76	5273901-11	Patent No. 5273901
20	7	3.2	76	5482709-10	Patent No. 5482709
21	7	3.2	76	5496550-6	Patent No. 5496550
22	7	3.2	79	US-09-252-991A-25964	Sequence 25964, A
23	7	3.2	118	US-08-301-162-10	Sequence 10, Appl.
24	7	3.2	118	US-09-461-240-10	Sequence 10, Appl.
25	7	3.2	118	US-09-968-927-10	Sequence 10, Appl.
26	7	3.2	122	US-09-328-352-7934	Sequence 7934, Ap
27	7	3.2	134	US-09-252-991A-18886	Sequence 18886, A

28	7	3.2	136	US-09-733-210-814	Sequence 814, App
29	7	3.2	136	US-09-733-210-815	Sequence 815, App
30	7	3.2	141	US-09-252-991A-26122	Sequence 26122, A
31	7	3.2	156	US-08-074-121-5	Sequence 5, Appl.
32	7	3.2	156	PCT-US94-06447-5	Sequence 5, Appl.
33	7	3.2	159	US-08-481-435-13	Sequence 13, Appl.
34	7	3.2	170	US-09-252-991A-27069	Sequence 27069, A
35	7	3.2	173	US-09-252-991A-17373	Sequence 17373, A
36	7	3.2	173	US-09-252-991A-25321	Sequence 25321, A
37	7	3.2	180	5273901-7	Patent No. 5273901
38	7	3.2	180	5482709-6	Patent No. 5482709
39	7	3.2	188	US-09-252-991A-20399	Sequence 20399, A
40	7	3.2	195	US-09-252-991A-20967	Sequence 20967, A
41	7	3.2	197	US-09-252-991A-24218	Sequence 24218, A
42	7	3.2	202	US-09-252-991A-28326	Sequence 28326, A
43	7	3.2	216	US-09-252-991A-18761	Sequence 18761, A
44	7	3.2	239	US-09-252-991A-21250	Sequence 21250, A
45	7	3.2	249	US-09-252-991A-18733	Sequence 18733, A

ALIGNMENTS

```

RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 631413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match      100.0%; Score 219; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.1e-203;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKLLIMACTIVCAFAKRRRPFPIGDDNDGHPHPSINIPYGINLBPPLYRRVNTVP 60
      |||
DB      1 MKLLIMACTIVCAFAKRRRPFPIGDDNDGHPHPSINIPYGINLBPPLYRRVNTVP 60
      |||

QY      61 SYPGNTYDTGGLPSYPIILTSRPGFPYVYHIGFPLATOLNVPPLPRGFPVPSRFSFA 120
      |||
DB      61 SYPGNTYDTGGLPSYPIILTSRPGFPYVYHIGFPLATOLNVPPLPRGFPVPSRFSFA 120
      |||

QY      121 AAAPAPPIAEPAAAPLTAIPVAABPAAGAPVAABPAEAPVAABPAEAPVAABPA 180
      |||
DB      121 AAAPAPPIAEPAAAPLTAIPVAABPAAGAPVAABPAEAPVAABPAEAPVAABPA 180
      |||

QY      161 EAPVGEPAABEPSPAEPAATKAPAPBPHPSLEQANO 219
      |||
DB      161 EAPVGEPAABEPSPAEPAATKAPAPBPHPSLEQANO 219
      |||

RESULT 2
US-08-428-414A-2
; Sequence 2, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; LEISHMANIASIS

```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-Apr-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecsek, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-414A-2

Query Match
Best Local Similarity 4.1%; Score 9; DB 2; Length 322;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAP 138
DB 289 AAEPAAAP 297

RESULT 3
US-09-198-452A-96
Sequence 96, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Grifais, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 96
LENGTH: 132
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-96

Query Match
Best Local Similarity 3.7%; Score 8; DB 4; Length 132;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAP 178
DB 53 EAPVAAP 60

RESULT 4
US-09-252-991A-18199
Sequence 18199, Application US/09252991A

Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18199
LENGTH: 168
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18199

Query Match
Best Local Similarity 3.7%; Score 8; DB 4; Length 168;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 VAAEPAG 151
DB 108 VAAEPAG 115

RESULT 5
US-09-252-991A-17963
Sequence 17963, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17963
LENGTH: 190
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17963

Query Match
Best Local Similarity 3.7%; Score 8; DB 4; Length 190;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAAP 128
DB 113 AAAPAAP 120

RESULT 6
US-09-252-991A-21588
Sequence 21588, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21588
LENGTH: 307
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21588

Query Match 3.7%; Score 8; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAP 128
Db 120 AAAPAP 127

RESULT 7
US-09-471-396-3
Sequence 3, Application US/09471396
Patent No. 6458359
GENERAL INFORMATION:
APPLICANT: BEDATE, Carlos Alonso
APPLICANT: REQUEÑA ROLANTA, Jose Maria
APPLICANT: SOTO ALVAREZ, Manuel
TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
FILE REFERENCE: bedate2a.seq
CURRENT APPLICATION NUMBER: US/09/471,396
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,825
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 383
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Protein Q
US-09-471-396-3

Query Match 3.7%; Score 8; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAP 127
Db 354 AAAPAP 361

RESULT 8
US-09-788-345-10
Sequence 10, Application US/09788345
Patent No. 6525186
GENERAL INFORMATION:
APPLICANT: ALONSO BEDATE, Carlos
APPLICANT: REQUEÑA ROLANTA, Jose M.
APPLICANT: SOTO ALVAREZ, Manuel
TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
FILE REFERENCE: BEDATE=1A
CURRENT APPLICATION NUMBER: US/09/788,345
CURRENT FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 09/219,306
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 383
TYPE: PRT

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chimeric
US-09-788-345-10

Query Match 3.7%; Score 8; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAP 127
Db 354 AAAPAP 361

RESULT 9
US-09-252-991A-30606
Sequence 30606, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30606
LENGTH: 410
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30606

Query Match 3.7%; Score 8; DB 4; Length 410;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156
Db 37 AAGAPVAA 44

RESULT 10
US-09-471-396-1
Sequence 1, Application US/09471396
Patent No. 6458359
GENERAL INFORMATION:
APPLICANT: BEDATE, Carlos Alonso
APPLICANT: REQUEÑA ROLANTA, Jose Maria
APPLICANT: SOTO ALVAREZ, Manuel
TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
FILE REFERENCE: bedate2a.seq
CURRENT APPLICATION NUMBER: US/09/471,396
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,825
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 412
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Protein Q
US-09-471-396-1

Query Match 3.7%; Score 8; DB 4; Length 412;
Best Local Similarity 100.0%; Pred. No. 8.5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
|||||

DB 383 AAAAPAP 390

RESULT 11
US-09-788-345-12

; Sequence 12, Application US/09788345

; Patent No. 6525186

; GENERAL INFORMATION:

; APPLICANT: ALONSO BEDATE, Carlos

; APPLICANT: ROQUELA ROLANTA, Jose M.

; APPLICANT: SOTO ALVAREZ, Manuel

; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN

; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG

; FILE REFERENCE: BEDATE-1A

; CURRENT APPLICATION NUMBER: US/09/788,345

; CURRENT FILING DATE: 2001-02-21

; PRIOR APPLICATION NUMBER: 09/219,306

; PRIOR FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 12

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: recombinant sequence from vectors p031 and pma1

Query Match
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
|||||

DB 383 AAAAPAP 390

RESULT 12

US-09-252-991A-20196

; Sequence 20196, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfeld et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20196

; LENGTH: 549

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-20196

Query Match
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 AAEPAPGA 152
|||||

DB 326 AAEPAPGA 333

RESULT 13
US-08-653-648A-5

; Sequence 5, Application US/08653648A

; Patent No. 6379945

; GENERAL INFORMATION:

; APPLICANT: Jepson, Ian

; APPLICANT: Martine, Andrew

; APPLICANT: Martinez, Alberto

; TITLE OF INVENTION: A Gene Switch

; FILE REFERENCE: PP050047/US

; CURRENT APPLICATION NUMBER: US/08/653,648A

; CURRENT FILING DATE: 1996-05-24

; PRIOR APPLICATION NUMBER: GB 9510759.5

; PRIOR FILING DATE: 1995-05-26

; PRIOR APPLICATION NUMBER: GB 9605656.9

; PRIOR FILING DATE: 1996-03-18

; PRIOR APPLICATION NUMBER: GB 9513882.2

; PRIOR FILING DATE: 1995-07-07

; PRIOR APPLICATION NUMBER: GB 9517316.7

; PRIOR FILING DATE: 1995-08-24

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 5

; LENGTH: 575

; TYPE: PRT

; ORGANISM: Helicobacter pylori

; US-08-653-648A-5

Query Match
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 TATPVAAE 147
|||||

DB 552 TATPVAAE 559

RESULT 14

US-09-252-991A-27591

; Sequence 27591, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfeld et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27591

; LENGTH: 778

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-27591

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 PAAGAPVA 155
|||||

DB 655 PAAGAPVA 662

RESULT 15

US-08-804-227C-14

; Sequence 14, Application US/08804227C

; Patent No. 5876991

; GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.
APPLICANT: Kutschos, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-14

Query Match 3.7%; Score 8; DB 2; Length 4545;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 AEPAAAP 138
+ |||||
Db 2620 AEPAAAP 2627

Search completed: January 15, 2004, 15:34:05
Job time : 21 secs

THIS PAGE BLANK (USPTO)